This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

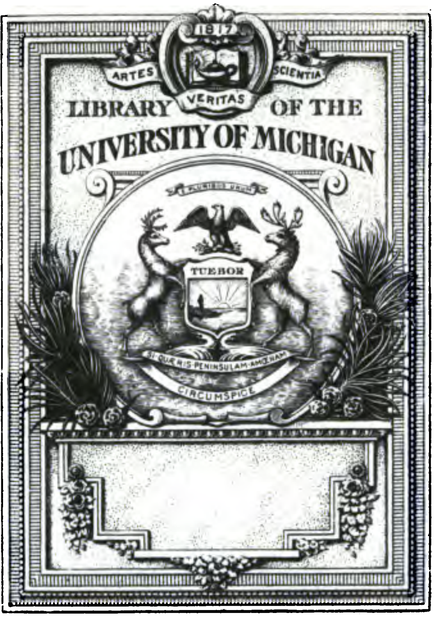
+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

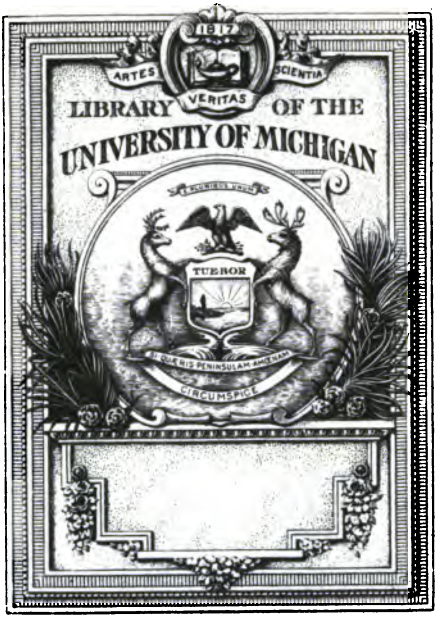
About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`



THE GIFT OF
Mrs. J. O. Reed

QC
21
.P37
1870



LIBRARY OF THE
UNIVERSITY OF MICHIGAN

THE GIFT OF
Mrs. J. O. Reed

QC
21
.P37
1870

A MANUAL OF PHYSICS.

A MANUAL OF PHYSICS

BEING

AN INTRODUCTION TO THE STUDY OF PHYSICAL SCIENCE

Designed for the Use of University Students and Advanced Classes in Secondary Schools.

BY

WILLIAM PEDDIE D.Sc., F.R.S.E.,

ASSISTANT TO THE PROFESSOR OF NATURAL PHILOSOPHY, AND LECTURER ON
ADVANCED EXPERIMENTAL PHYSICS, IN THE UNIVERSITY
OF EDINBURGH.

SECOND EDITION.

(REVISED AND ENLARGED.)

New York: G. P. PUTNAM'S SONS.
London: BAILLIÈRE, TINDALL & COX.
1896.

Gift
Mrs J.O. Reed
4-11-33

4-15-33 MEW

PREFACE TO THE SECOND EDITION.

In preparing this edition my chief aim has been to make the book more suitable to the wants of junior students. With this object, I have entirely rewritten the mathematical portions, and have used none but those elementary methods which may be readily followed by any intelligent schoolboy. The great objection to the use of the simpler methods is the special nature of the proofs which are used in different cases. I have endeavoured to avoid the want of unity which is apt to result from this.

These changes have been introduced with the view of making the book appeal, on the whole, to a wider class of readers. In order that it may still retain whatever advantages it may formerly have possessed for more advanced students, I have taken special care to obtain simple elementary proofs even in those cases which are invariably treated, so far as I know, by higher methods. Thus I have never deleted any portion merely because it was difficult, but have endeavoured rather to make it simple. And many portions, which were previously treated briefly, have been greatly expanded.

In a new chapter on the passage of electricity through gases, the treatment adopted by J. J. Thomson in his 'Recent Researches in Electricity and Magnetism' has been followed. The former treatment of the thermo-dynamical relations has been entirely altered, the subject being now discussed as a special case in a new chapter on related physical properties. Under this heading, one of the most interesting branches of physical science, generally treated only by very advanced methods, is dealt with, and has been brought, I trust, within the reach of the ordinary student. The arrangement of the chapters on magnetism and electromagnetism has been entirely altered. Electromagnetic phenomena have been treated directly as phenomena accompanying the flow of electricity, and magnetism has then been introduced through the consideration of the influence of the medium.

Briefly, the book is intended to be an accompaniment to a course of lectures on physics, by the use of which a student may largely avoid the evils of note-taking and give more attention to the words of his teacher. Even without such a course, it may perhaps be profitably used in conjunction with one of the published collections of physical problems and examples.

I have again to record my indebtedness to Mr. J. B. Clark for much help and many valuable suggestions. As an example of these I may refer specially to the rearrangement of the chapters on electromagnetism and magnetism, already spoken of. I have also to thank Mr. A. Alexander, Fraser Bursar in the Physical Laboratory, for his very careful revision of the proofs.

Lastly, I have to thank those teachers and students who have made use of the book, and to express the hope that the present edition may be found to be still more suited to their needs.

WILLIAM PEDDIE.

EDINBURGH UNIVERSITY,
November, 1895.

PREFACE TO THE FIRST EDITION.

THE best advice which can be given to a student of physics regarding the books which he should read is to use separate works written on the various branches of the subject by the leading physicists of the day. Yet this advice has one evident disadvantage. The student who follows it may not get so complete a view of the essential unity and interdependence of the various branches of his subject as it is desirable that he should. And, besides this, there is no doubt that a small volume which gives, as far as is possible, a review of the elements of the whole subject, is a desideratum to the student while in attendance on University classes. I have undertaken the writing of this work in the hope that it may to some extent meet that want.

I have throughout endeavoured to bring into prominence the necessity for, and the value of, scientific hypotheses—a matter regarding which very hazy notions are only too common.

It has also been my aim to make the treatment of the mathematical portions of the text as simple as possible. In this connection I have not adopted the process which has recently been termed 'calculus-dodging,' for the reason that the elementary methods of the calculus are more simple, certainly are more natural, than the methods by which they are usually supplanted.

At the same time it may be well to remark that any student, who desires to do so, may simply assume the results of the mathematical portions, and use the remainder (which is much the larger part) of the text in his study of experimental physics.

In writing a text-book on general physics it is impossible, if justice is to be done to the subject, to avoid borrowing methods from the writings of the masters. In this respect I have to acknowledge my indebtedness to the works of v. Helmholtz, Clerk-Maxwell, Thomson, Tait, and others. This is perhaps most evident in Chapter XXVI., where I have made use of the very simple form in which Tait has presented the analytical treatment of the theory of Thermodynamics; and in Chapter XI., where I have adopted his mode of discussing the compressibility and rigidity of solids. This apart, I have endeavoured, whether successfully or not, to present the various subjects in as fresh a manner as I could.

My indebtedness to Professor Tait has also been very great in the matter of criticism, which he kindly afforded me on various points while the book was passing through the press. I have also to acknowledge with thanks the kindness of Mr. J. B. Clark, M.A., F.R.S.E., Physical Master in Heriot's Hospital, Edinburgh, in reading the work both in manuscript and in proof. His careful revision has resulted in the elimination of a number of defects which had escaped my notice.

WILLIAM PEDDIE.

EDINBURGH UNIVERSITY,
November, 1891.

CONTENTS.

CHAPTER I.

INTRODUCTORY: THE PHYSICAL UNIVERSE.

PAGE

Physical science.—Test of reality: conservation.—Matter: its conservation: its inertia.—Energy: its transformation: its conservation: its degradation . . . . 1—8

CHAPTER II.

THE METHODS OF PHYSICAL SCIENCE.

Observation and experiment.—Cause and effect.—Hypothesis and theory.—Crucial test.—Argument from analogy.—Principle of stable equilibrium.—Errors of observation.—Instrumental errors.—Empirical laws.—The graphical method . . . . 9—16

CHAPTER III.

THE GRAPHICAL METHOD IN PHYSICS.

Dimensions.—Contour lines.—Special physical applications. 17—22

CHAPTER IV.

MOTION.

Position.—Displacement.—Speed and velocity.—Acceleration of speed.—Motion in a vertical line under gravity.—Resolution and composition of velocities and accelerations.—Curvature.—Acceleration of velocity.—Motion of projectiles.—The hodograph.—Moments.—Simple harmonic motion.—Composition of simple harmonic motions.—

PAGE

Wave motion along a line.—Rotation.—Uniplanar motion of a rigid body.—Motion of a rigid body in space.—Composition of angular velocities.—Displacement of the parts of a non-rigid body.—Strain.—Homogeneous strain.—Shear.—Non-homogeneous strain.—Motion of fluids.—Fluid circulation.—Vortex motion . . . 23—50

CHAPTER V.

MATTER IN MOTION.

Force.—Friction.—The laws of motion.—The first and second laws.—Examples.—Dynamical similarity.— The third law.—Impact.—Centre of inertia.—Moment of a force and of inertia.—Rotational equilibrium.—Propagation of motion through a non-rigid solid.—Motion of a perfect fluid.—Equilibrium of a fluid.—Propagation of waves in liquids 51—71

CHAPTER VI.

PROPERTIES OF MATTER.

Definitions of matter.—States of matter.—General properties.—Special properties.—Contingent properties.—Specific properties.—Molecules and atoms.—Molecular forces.—Gravitation.—Weight and mass.—Specific gravity.—Kepler's laws with Newton's deductions.—The law of gravitation.—Newton's theorems regarding spherical shells.—Mean density of the earth.—Hypotheses framed to explain gravitation.—The nebular hypothesis . . 72—86

CHAPTER VII.

PROPERTIES OF GASES.

Compressibility.—Boyle's law.—Compressibility of a perfect gas.—Deviations from Boyle's law.—Compression of vapours.— Elasticity.—Viscosity.—Diffusion.—Effusion.—Transpiration.—Atmolysis . . . . . 87—96

CHAPTER VIII.

PROPERTIES OF LIQUIDS.

Compressibility.—Elasticity.—Viscosity and viscidity.— Diffusion.—Osmose.—Dialysis.—Cohesion.—Capillarity.—Surface-tension . . . . . . 97—10

CHAPTER IX.

PROPERTIES OF SOLIDS.

PAGE

Compressibility and rigidity.—Elasticity.—Viscosity.—Cohesion.—Tenacity . . . . . . 108—121

CHAPTER X.

THE CONSTITUTION OF MATTER.

The hard atoms of Lucretius.—Boscovich's centres of force.—Vortex atoms.—Heterogeneity of matter.—Structure of crystals.—Size of molecules and the range of molecular forces . . . . 122—136

CHAPTER XI.

THE KINETIC THEORY OF MATTER.

Kinetic theory of gases.—Gaseous pressure.—Boyle's law.—Avogadro's and Charles' laws.—Diffusion.—Thermal conductivity.—Viscosity.—Evaporation.—Dissociation, etc.—Difficulties of the theory.—General kinetic theory of matter . . . . 137—148

CHAPTER XII.

SOUND.

Propagation.—Intensity.—Reflection.—Refraction.—Interference.—Diffraction.—Pitch.—Musical intervals.—Vibrations of rods.—Vibrations of plates.—Vibrations of strings and membranes.—Vibrations of air-columns.—Speed of sound in a gas.—Partial tones.—Resonance.—Quality.—Beats.—Dissonance.—Beat tones.—Combination tones.—Dissonance of pure tones . . . . . 149—176

CHAPTER XIII.

LIGHT: INTENSITY, SPEED, THEORIES, COLOUR.

Rectilinear propagation.—Intensity.—Speed.—Theories.—Colour.—Colour-mixing.—Vision.—Colour-blindness.—Complementary colours.—Contrast colours.—Young-Helmholtz theory.—Hering's theory . . . . 177—197

CHAPTER XIV.

LIGHT: REFLECTION, REFRACTION, DISPERSION.

PAGE

Laws of reflection.—Reflection from plane surfaces.—Reflection from curved surfaces.—Reflection caustics.—Focal lines.—Theoretical deductions of the laws of reflection.—Laws of refraction.—Refraction through a plane surface.—Refraction caustics.—Focal lines.—Refraction through parallel layers.—Mirage.—Prisms.—Refraction through spherical surfaces.—Lenses.—Oblique refraction.—Formation of images by lenses.—Dispersion.—Aberration.—Achromatism.—Rainbows.—Halos.—Theoretical deductions of the laws of refraction . . . . . 198—229

CHAPTER XV.

RADIATION AND ABSORPTION: SPECTRUM ANALYSIS.—ANOMALOUS DISPERSION.—FLUORESCENCE.

Equality of emissivity and absorptive power.—Spectrum analysis.—Law of absorption.—Body colour.—Dichroism.—Surface colour.—Metallic reflection.—Anomalous dispersion.—Fluorescence.—Phosphorescence.—Theories of dispersion . . . . 230—246

CHAPTER XVI.

INTERFERENCE.—DIFFRACTION.

Principle of interference.—Young's experiment.—Fresnel's experiment.—Lloyd's experiment.—Fresnel's biprism.—Coloured interference bands.—Michelson's method.—Displacement of bands by refracting media.—Interference bands in spectra.—Colours of thin plates.—Newton's rings.—Colours of mixed plates.—Colours of thick plates.—Diffraction.—Effect of a rectilinear wave.—Effects of plane and spherical waves.—Diffraction at a straight edge.—Diffraction at a narrow slit.—Diffraction at a circular aperture.—Zone plates.—Diffraction at an opaque disc.—Coronæ.—Young's Eriometer.—Diffraction gratings 247—271

CHAPTER XVII.

DOUBLE REFRACTION.—POLARISATION.

Double refraction.—Huyghens' construction.—Special sections of the surface.—Polarisation.—Laws of polarisation by

PAGE

reflection and refraction.—Direction of vibration in polarised light.—Reflection and refraction of polarised light.—Plane, circular, and elliptic polarisation.—Nature of common light.—Metallic reflection.—Double refraction in biaxal crystals.—Fresnel's theory of double refraction.—Conical refraction.—Interference of polarised light.—Colours of crystalline plates.—Haidinger's brushes.—Artificial production of the doubly refracting structure.—Rotatory polarisation.—Polarising prisms . . 272—302

CHAPTER XVIII.

THE NATURE OF HEAT.

Radiant heat.—Its identity with light.—Heat in material bodies.—Hypothesis of molecular vortices . . 303—307

CHAPTER XIX.

RADIATION AND ABSORPTION OF HEAT.

Prevost's theory of exchanges.—Stewart's and Kirchhoff's extension of Prevost's theory.—Laws of radiation of heat at constant temperature.—Heat spectra.—Diathermancy.—Radiation at different temperatures.—Solar radiation.—Radiation from moving bodies . . . 308—318

CHAPTER XX.

EFFECTS OF THE ABSORPTION OF HEAT: DILATATION.

Temperature.—Dilatation of solids.—Dilatation of liquids.—Dilatation of gases.—Absolute zero of temperature.—Measurement of temperature . . . 319—333

CHAPTER XXI.

EFFECTS OF THE ABSORPTION OF HEAT: CHANGE OF TEMPERATURE AND CHANGE OF STATE.

Unit of heat.—Specific heat.—Thermal capacity.—Specific heat of solids and liquids.—Law of Dulong and Petit.—Specific heat of gases and vapours.—Change of molecular state.—Latent heat.—Fusion and solidification.—Latent heat of fusion.—Evaporation and condensation.—Latent heat of vaporisation.—Formation of dew.—Continuity of the liquid and gaseous states.—Critical temperature.—

PAGE

Solution.—Freezing mixtures.—Dissociation and chemical combination . . . . 334—356

CHAPTER XXII.

CONDUCTION AND CONVECTION OF HEAT.

Conduction.—Conductivity.—Measurement of conductivity.—Conduction through the earth's crust.—Conduction in crystalline bodies.—Conduction in liquids and gases.—Convection . . . . 357—369

CHAPTER XXIII.

THERMODYNAMICS: HEAT AND WORK.

Mechanical equivalent of heat.—First law of thermodynamics.—Carnot's complete cycle of operations.—Carnot's reversible cycle.—Reversibility the test of perfection.—Second law of thermodynamics.—Absolute temperature.—The indicator diagram and its applications.—Entropy.—Total, available, and dissipated energy.—Thermodynamic motivity . . . . 370—385

CHAPTER XXIV.

ELECTROSTATICS.

Electrification by friction.—Conductors and non-conductors.—Fundamental phenomena presented by electrified bodies.—Positive and negative electricity.—The gold-leaf electroscope.—Electric charge and quantity.—Charging by induction.—Continued production of electricity.—The electrophorus.—Law of electric attraction and repulsion.—Electric potential.—Electromotive force.—Capacity.—Condensers.—Specific inductive capacity.—Distribution of electricity on conductors.—Electric density.—Electric images.—Equipotential surfaces.—Lines and tubes of force.—Total force over a closed surface.—Special applications.—Electric induction.—Tubes of induction.—Electric energy.—Electric absorption.—Disruptive discharge.—Atmospheric electricity, etc.—Pyroelectricity.—Peizo-electricity.—Electrification by contact.—The electrometer.—Electric machines . . . . 386—431

CHAPTER XXV.

THERMO-ELECTRICITY.

PAGE
Thermo-electric phenomena.—Laws of thermo-electric circuits.—Variation of the electromotive force with temperature.—The thermo-electric diagram.—The Peltier effect.—The Thomson effect.—Further discussion of the thermo-electric diagram.—The thermopile and radiomicrometer . 432—443

CHAPTER XXVI.

ELECTRIC CURRENTS.

Convection current between charged conductors.—Flow of electricity in metallic conductors.—Ohm's law.—Kirchhoff's laws.—Conductivity and resistivity.—Electrolytic conduction.—Faraday's laws.—Polarisation.—Von Helmholtz's electrochemical theory.—Ideas of Grotthus and Clausius.—Ohm's law in electrolytes.—Production of electric currents.—Primary cells.—Voltaic batteries.—Secondary cells.—Transformations of electric energy in conducting circuits.—Joule's law.—Measurement of electromotive force, current strength, and resistance.—Conductance expressed as a velocity.—Currents of varying strength . 444—465

CHAPTER XXVII.

ELECTRIC CURRENTS IN GASES.

Molecular or atomic charge.—Effect of violet light.—Effect of glowing metals.—Electric strength, difference of potential, and spark discharge.—Spark length and potential difference in a practically uniform field.—Spark length and potential difference in a non-uniform field.—Maximum electromotive intensity and spark length.—Spark potential and gaseous pressure.—The electrodeless discharge.—Molecular conductivity of gases.—Passage of electricity between dissimilar substances.—Discharge with electrodes.—The negative rays.—Variation of potential from anode to cathode.—Variations of pressure during discharge.—The arc discharge.—The discharge in a magnetic field.—Chemical effects.—General conclusions . . 466—480

CHAPTER XXVIII.

ELECTRO-MAGNETISM.—MAGNETISM.

PAGE

Magnetic effects of electric currents.—Magnetic shell.—Intensity of magnetisation.—The law of magnetic attraction and repulsion.—Magnetic potential —Magnetic potential due to a shell.—Circular circuits.—Solenoids.—Influence of the medium.—Magnetic and diamagnetic media.—Material magnets.—Magnetism a molecular phenomenon in material media.—Experimental mapping of lines of magnetic force. —Magnetic induction.—Permeability and susceptibility.—Law of magnetic circuits.—Relation between magnetisation and magnetising force.—Hysterēsis.—Effects of vibration and of temperature.—Effects of stress.—Magnetometric measurements.—Terrestrial magnetism.—Theories of magnetism.—Ampère's hypothesis . . . . 481—509

CHAPTER XXIX.

ELECTRO-MAGNETIC INDUCTION.—ELECTRO-DYNAMICS.

Action of an external magnetic field on an electric circuit.—Induced electromotive force.—Law of rise and decay of currents.—Special cases of induced currents.—Electrokinetic energy.—The ballistic method.—The simple harmonic alternating current.—Law of conduction in an inductive circuit.—Law of conduction in a condensing circuit.—Total conduction in a condensing and inductive circuit.—Alternate current transformers and electromagnets.—Alternate current dynamos.—Continuous current dynamos.—Characteristic curves of dynamos.—Constant potential, and constant current, dynamos.—Electro-dynamic action on an electric circuit.—Case of linear circuits.—Continuous motion under electro-dynamic action.—Electric motors.—Electrical efficiency.—Measurement of alternating currents and electro-motive forces.—Measurement of power.—Resistance expressed as a velocity. —Electric and magnetic units.—Practical units.—Explanation of the phenomena of electric and magnetic fields by motion of Faraday tubes.—Historical development of the subjects . . . . 510—535

CHAPTER XXX.

ELECTRO-MAGNETIC THEORY OF LIGHT.

PAGE

Magnetic rotation of the plane of polarisation of light.—Hypothesis of molecular vortices.—Hall's effects.—Kerr's effects.—Photo-electric actions.—Electro-magnetic theory of light.—Electro-magnetic waves.—Waves in wires 536—547

CHAPTER XXXI.

RELATED PHYSICAL PROPERTIES.

General law of transformation of energy.—Absolute value of the potential factor.—General relations.—Heat and mechanical work.—Latent heat and work during evaporation.—Heat and electrical energy.—Electrical energy and work.—Electrification and vapour pressure.—Surface tension and vapour pressure.—Induced electromotive forces.—Surface tension and solubility.—Pressure and solubility.—Relation between the specific heats of a perfect gas.—Determination of the absolute scale of temperature.—Basis of the general law of transformation . . 548—557

CHAPTER XXXII.

THE ETHER.

Direct action at a distance.—Action through a medium.—Rigidity and density of the ether.—Elastic-solid theory.—Contractile ether.—Electro-magnetic ether.—Larmor's theory.—Dilatancy . . . . . . 558—565

CHAPTER XXXIII.

THE BASIS OF PHYSICAL BELIEFS, AND THE ADVANCEMENT OF PHYSICAL SCIENCE 566, 567

INDEX . . . . 569—573

ERRATUM.

§ 347. Delete '*Ampère's Hypothesis regarding Magnetism.*'

A MANUAL OF PHYSICS.

CHAPTER I.

THE PHYSICAL UNIVERSE.

1. All processes which occur in the universe may be classified as physical or as non-physical. They appertain essentially on the one hand to dead matter, or, on the other hand, to matter which possesses (or is possessed by) life. This statement, of course, is a mere definition of what is meant by the word *physical*, and is not to be regarded as being in any sense the expression of an opinion regarding the nature of life. If in the phenomena associated with living bodies we meet with processes akin to others which occur in the inorganic world, we regard them as being purely physical; while if in these phenomena we meet with processes which are totally dissimilar to any with which we are acquainted in the phenomena of dead matter, we leave their further study to the biologist.

2. It is evident, therefore, that the domain of the physical sciences is of immense extent—of such extent that no one can hope, in the course of the longest life, to master all its known details. At one time, indeed (and that not very remote), the scientist might have said, after the fashion of Francis Bacon, 'I have taken all knowledge for my realm,' but such a claim would be impossible now. All questions regarding the combinations and interactions of the various kinds of matter are purely physical in their nature, although their study is now left to specialists in the department of chemistry. The investigation and prediction of the motions of the heavenly bodies and the determination of their physical constitution are left to the astronomer, while the configuration of the earth is studied by the geographer. So also the sciences, as distinguished

from the arts, of navigation and ship-designing, the sciences of engineering, mineralogy, geology (in large part, at least), meteorology, and so on, are purely physical sciences, the study of which is undertaken by specialists, for each of them is more than a life-study for any one man. Thus with increase of knowledge the domain of the physical sciences has been subdivided, and the equivalent terms, *Physical Science*, *Natural Philosophy*, *Physics*, have become restricted in meaning, so that they refer merely to the pure scientific groundwork which underlies all the more practical or more highly specialized physical sciences.

3. In commencing the study of physics we have no concern with purely metaphysical questions regarding the objective reality or non-reality of the universe. We simply assume that it has an existence quite independently of the existence of an observing mind, and then proceed to examine the facts and phenomena which make up its entirety. But, before entering upon any detailed examination, it is well that we should take a glance at our subject as a whole, in order to learn something of its scope and of the mutual relations of its various parts; and this just for a similar reason to that which makes it desirable for the traveller in an unknown country to examine it first from the vantage-ground of some commanding height, so that he may carry with him in his future wanderings therein a clear mental picture of its disposition.

4. A most obvious difficulty meets us at the very outset. How are we to distinguish between that which has true existence and that which has only the appearance of it—between the true landscape, as it were, and the mirage? The mirage seems to the observer to be as real as the reality. What, then, is to be the test of true existence?

In addition to the assumption of the real existence of the physical universe, it is assumed that in this universe *there is nothing which does not occur according to law*. But this assumption is not left without support. The mere possibility of the existence of the so-called exact sciences may be taken as evidence of its truth. Keeping this idea in view, we see that no real thing can appear in, or disappear from, the universe in an arbitrary manner. If it appears or disappears in accordance with some law, we admit that it *may* be real; but we can only with certainty regard it as a reality when we can prove it to be constant in amount—that is (to use the ordinary scientific expression), when we can show that it possesses the property of *conservation*. CONSERVATION is our great test of reality.

This test being applied, it is found that there are two, and only two, classes of things in the physical world—MATTER and ENERGY—which sustain it.

5. Everyone knows what is meant by the term the 'matter' or 'material' of which a body is composed, so, in the meantime, no attempt at a definition is necessary.

At first sight it might appear that matter is certainly *not* conserved. If we weigh a lump of pure limestone (carbonate of lime), so as to determine the amount of matter in it, and then heat it sufficiently, we find that its weight has become less during the process. It would seem that matter has been lost. But what has actually occurred is the decomposition of the carbonate of lime, by the application of the heat, into lime and carbonic acid gas. The latter, being colourless, passes off unnoticed, and the second weighing gives only the weight of the lime. By proper means the weight of the gas may be determined, and it is then found that the original weight of the limestone is equal to the sum of the weights of its constituents. That particular kind of matter called limestone has been altered in amount in accordance with a definite law; but no matter as a whole has been lost in the process. And in *all* chemical processes, however complex, the same result holds; indeed, the science of chemistry is a possibility only in virtue of the strict conservation of matter. Therefore we say that matter is a real thing.

There are many kinds of matter which differ from one another to a greater or less extent in their various physical properties. These properties will be considered in detail in subsequent chapters, and their variations from one to another of a few of the most important or most peculiar substances will be indicated; but the enumeration, classification, and investigation of the various substances in nature belong more to the science of chemistry than to that of physics. One special substance, which pervades all others, and extends throughout the whole of the visible universe, possesses such extraordinary properties, and is of such immense importance physically, that it must receive separate treatment (Chap. XXXII.).

6. In addition to the property of conservation, matter is characterized by passivity or inertness. It is said to possess INERTIA. In other words, a material body can do nothing of itself. If at rest, it cannot move unless something outside of itself sets it in motion. If moving, it cannot come to rest or alter its motion in any way unless something external to it produce that effect. This property and its consequences will be discussed further in Chap. V. It is the distinguishing characteristic of matter.

7. As in the case of matter, so in the case of energy, no formal

definition of the term is necessary or desirable. There is every reason to believe that the inertia of matter cannot be overcome; that is, that the motion of matter cannot be altered, unless motion is imparted to it from some other portion of matter, whether by direct collision or otherwise. It is commonly said in such a case that the second body 'does work' upon the first, or that the first has 'work done' upon it by the second. Thus 'the doing of work' involves essentially the production of motion.

It is customary to say that the body to which motion has been imparted has acquired ENERGY, the meaning being that something is now associated with the matter of the body in virtue of which it can do more work than formerly upon some other body. The measure of the energy acquired is the amount of work which has been done upon the body, and the measure of the total amount of energy which it possesses is the total amount of work which it can do. In no case are we able experimentally to determine the total amount, but in no case is it necessary that we should do so; we only require to know its variations.

Although in all likelihood the possession of work or energy necessarily involves motion of the material system which possesses it, the moving system may not be evident to our senses, in which case it is convenient to speak of the energy as existing *potentially* in some connected system which may be at rest at the time, but which can be set in motion by having energy imparted to it from the other. (See next section.)

It is easy to determine the increase of energy of a body, originally at rest, which is made to move with a given speed. Of two bodies moving with a common speed, that one with least mass (quantity of matter) can do least work, and the same is true of the slower moving of two bodies which have the same mass. The work tends to vanish either as the mass or the speed becomes indefinitely small; and so far as experiment (which is the only permissible test) shows, it does not depend upon anything else than the mass and the speed.

Each unit of mass, moving with a given speed, is found to possess the same amount of energy; and, therefore, the energy of any body is proportional to the total quantity of matter which it contains.

Now suppose that two bodies, whose masses are equal in amount, are moving in the same straight line towards each other with equal speeds. Each possesses an equal amount of energy, E (say). We may assume that the result of the impact is that each body is brought to rest, so that the quantity of energy $2E$ has been expended in stopping the forward motion of the two masses. Next, let one of the bodies be at rest while the speed of the other is doubled. The relative speed of the two is unaltered, and hence the energy expended in impact is still $2E$. But experiment shows that the two bodies will now move together with half the speed of the single one, so that each body still possesses energy E. Therefore the single body, moving with double speed, possesses an amount of energy $4E$.

By this experiment, and by other similar experiments conducted under varying conditions regarding the speeds of the moving bodies, it may be proved that the energy of a moving body varies directly as the square of its speed.

If E be the energy of the moving body, we may express the results just obtained by means of the equation

$$E=kmv^2,$$

where k is a constant. The value of k is purely arbitrary, depending on our choice of the unit of energy, but it will be shown subsequently that it is convenient to choose the units, so that the value of k is $\frac{1}{2}$. Thus

$$E=\tfrac{1}{2}mv^2.$$

8. Energy, when it is regarded as in the preceding paragraph, is called energy of motion or *kinetic energy*. But, as already indicated, we do not always perceive the system to which such energy is communicated. The kinetic energy which the visible system still possesses becomes less and less, and at last, in a certain position of the system, it may entirely vanish. The kinetic energy remaining at any instant may obviously be expressed in terms of the position of the system, and so also may the energy which has been given away from it. Therefore, on the assumption, made in § 7, that the gain of energy of the invisible system is equal to the loss of energy of the visible one, we can represent the kinetic energy of the invisible system in terms of the position of the visible one. This is always done when the two systems are so connected that, when left to themselves, the energy will be recommunicated to the visible system, and so we speak of energy of position or *potential energy*.

As a special example we may consider the case of a bullet fired vertically upwards. The further it rises the less its speed becomes, and the less work it can do in overcoming obstacles. At last it comes to rest, and is totally devoid of kinetic energy. But we have only to let it fall down again, and (neglecting the resistance of the air, which only complicates the problem without altering its essence) we find that its energy of motion, when it again reaches the ground, has the same value as at first. Hence, instead of saying that as the ball loses energy some connected system gains it, we, for convenience, say that as it loses kinetic energy it gains potential energy.

We do not yet know what this connected system, in the case of gravitation, is. If Le Sage's hypothesis of ultra-mundane corpuscles (Chap. VI.) were true, the kinetic energy of a ball, projected upwards against gravity, would be transmuted into kinetic energy of the corpuscles.

We may distinguish a number of forms of energy, all of which can be classified under the two main types just defined. Kinetic and potential energy of visible portions of matter have been already considered. There is also kinetic energy of invisible portions of matter (Chap. XVIII.), as in the case of a body which is sensibly hot. And potential energy also exists on a similar scale, as in the case of the so-called Latent Heat (Chap. XXI.). Other examples appear in the molecular motion of gases, and in the transmission of vibrations through an elastic medium, etc. Again, two oppositely electrified bodies attract each other, and so have potential energy of electrical separation. And, when electricity flows along a conductor, the energy of electricity in motion becomes evident. Also two chemical substances which tend to combine and form a compound substance are said to have potential energy of chemical separation. Lastly, two magnets have potential energy relatively to each other, and work can be indirectly produced by means of the motion of magnets.

9. At this point we are led to regard that characteristic in the possession of which energy differs totally and fundamentally from matter. While matter is essentially passive, energy is constantly in a state of change. It is constantly being handed on from one portion of matter to another, and is ever being changed from one to another of the forms above indicated. It is said to possess the property of TRANSFORMATION.

Only in virtue of this property can we recognise its existence. We could never have known that a moving cannon ball possessed energy had we never seen its destructive effects. We would have been ignorant of the energy of an electrified thunder-cloud if we had not seen the production from it of light, heat, sound, and mechanical effect. *How* energy is passed on from one material system to another, and *how* it changes from one form to another, are questions to which no final answer can at present be given.

A simple example of the transformation of energy is furnished by the motion of an ordinary pendulum. At the lowest part of the swing the energy is entirely kinetic; at the highest part it is entirely potential; and, in intermediate positions, it is partly kinetic, partly potential.

A somewhat more complex example occurs in the transmission of a message by telephone. There is first the energy of vibratory motion of the air when the sound is produced. This vibratory motion is communicated to the metallic diaphragm of the telephone. But the diaphragm is magnetized by induction, and so its motion causes alterations in the intensity of magnetization of the magnet. These alterations in the magnetization produce electric currents in the wire coiled round the magnet, and these currents produce similar alterations of magnetization in the magnet of the receiving telephone, and so similar motions of its diaphragm ensue; consequently, similar sounds are heard at the receiving instrument.

By the consideration of such special examples we are led to the conclusion that any form of energy may, directly or indirectly, be changed into any other. Many evidences of this will appear in subsequent chapters.

10. As we have already learned, during all its changes and transferences one thing is evident regarding the energy in the universe — the total amount of it is unalterable. This is made clear by the fact that strict 'mechanical equivalents' of heat and the various other forms of energy are obtainable (Chap. XXIII.). Energy, like matter, possesses the property of CONSERVATION. The swings of a pendulum which has been set in motion and then left to itself gradually die away, and finally vanish. But if no energy were lost because of the communication of motion to the air, and if none were lost because of friction at the points of support, or because of vibrations set up in the supporting framework, etc., the motion would go on for ever; that is to say, the energy communicated to other bodies up to any instant, together with the energy still possessed by the pendulum at that instant, is equal in amount to the original quantity. And the same is true of any other system. Therefore, since it is conserved, we must regard energy as having real existence.

11. A question of the deepest importance to mankind arises in connection with the transformation of energy. Are all forms of energy equally transformable? When energy is changed from one form to another, can it with equal readiness be changed back again into the original form? If not, it necessarily follows that the whole amount of energy in the universe will gradually assume that particular form which is least transformable. Observation and experiment have shown that there is one form into which all others are gradually and permanently changing; and that form is the energy of molecular motion known as *heat*. But there is a constant

tendency towards diffusion of heat, so as to produce uniformity of temperature; and when uniformity of temperature is arrived at, no mechanical work can be produced from the heat. The total amount of energy in the universe will be the same as before, in accordance with the principle of conservation; but none of it will be available for the production of mechanical work. This principle of the loss of availability of energy is technically known as the principle of DISSIPATION, or (preferably perhaps) DEGRADATION of energy.

Examples of degradation of energy occur everywhere in nature. No stone falls from a cliff, no storm arises or ceases, no flash of lightning or peal of thunder occurs, no wave breaks upon the shore, without a diminution of the possible amount of useful work obtainable for man by natural processes.

In accordance with this principle, potential energy of visible portions of matter tends towards a minimum value—*i.e.*, tends as far as possible to take the form of kinetic energy. But further consideration of this subject must be deferred in the meantime.

12. Nothing which is not either matter or energy is conserved—at least, in the same sense as that in which we assert conservation of these things. In our consideration of matter and energy, we regarded both as signless quantities. We might assert conservation of a quantity which may be positive or negative, provided that, when a new positive amount of it is produced, an equal negative amount necessarily appears. In this case the total algebraic sum of the quantity is constant. But if we regard either the positive portion of it or the negative portion of it, we may find that the amount of either portion is variable at will: and this is not the sense in which conservation is asserted of matter and energy.

In the new sense alluded to, we speak of the conservation of momentum (Chap. V.), and sometimes even of the conservation of electricity.

CHAPTER II.

THE METHODS OF PHYSICAL SCIENCE.

13. THE whole body of scientific knowledge has been obtained by one or other of two methods—*observation* or *experiment:* nor can strict knowledge be obtained in any other way. The first scientific investigations ever made must have been of a purely observational type, and were very probably astronomical in their nature. In making observations, we notice the positions of objects and the sequence of events; and we attempt, then, to make out relations among them. In this way arose the still-extant grouping of the stars into various constellations, and — greatest perhaps of all examples—the discovery by Kepler of the laws which regulate the motions of the planets. But, when we alter at will the conditions attending certain phenomena, so as to discover the consequent alterations produced in the phenomena, we are said to experiment. It is true, indeed, that we cannot always draw a hard and fast distinction between observation and experiment. Thus, in calculating the speed of light from observations upon the satellites of Jupiter, although we do not ourselves alter any conditions, yet we purposely take advantage of alterations which occur naturally.

By such means we first of all obtain mere series of facts often without any mutual connection whatsoever, and, not infrequently, so grouped as to suggest false relations. The next duty of the scientist is to group these isolated data after a definite system, to co-ordinate the facts with the object of subsequently discovering the true relations which subsist among them; and the greater the power of the observer to detect real resemblances and essential differences the sooner will his ulterior object be attained.

The question of cause and effect next arises. Of two phenomena which appear successively, and no one of which appears without the other, that one which is first evident is usually called the cause of the other, which is said to be its effect. But, obviously, great care must be taken to avoid any error in such an assertion, for there may be many sources of mistake. In the first place, it is conceivable that two phenomena might appear in invariable succession the one to the other, and yet the true explanation might be that they had a common cause, and were not otherwise connected. The flash of forked lightning and the sound of thunder occur successively; but the sound is due to the explosive expansion of the air through which the electricity passes, while a portion of the light is also caused by this explosive expansion, which compresses the adjacent layers so suddenly as to render them luminous by the excessive heat developed. Again, the occurrence of one event is frequently necessary, in order that we may perceive another between which and the former there is no connection whatsoever; and, frequently, the effect becomes evident before the cause is noticed. Still further, we observe that events sometimes occur simultaneously. Thus, tornadoes are often due to the sudden heating of large portions of the atmosphere by means of the latent heat given out on rapid condensation of vapour. But the condensation of vapour and the evolution of latent heat occur of necessity at one and the same instant, so that we might with equal propriety refer the tornado to either event as a cause.

In nature there is an apparently endless series of causes. Each event is the cause of another, and was itself produced as the consequence of a preceding event. When a large mass of cloud intercepts the rays of the sun from the underlying atmosphere, the air grows colder, and as it grows colder it contracts. This causes an inrush of air from the surrounding regions, which well-known result is expressed by the popular phrase that the cloud or the rain 'draws' the wind. The effects of this motion might be traced out endlessly if our senses were sufficiently acute and our powers sufficiently universal; and so also the various motions preceding the motion of the cloud might be traced.

14. Among his other duties, the physicist has to undertake the investigation of the effects which result from physical conditions. Such an investigation is comparatively simple. He has only to make certain that the effects which he observes are not due to any unnoticed conditions. The converse problem—the investigation of causes—is not by any means so simple. The investigator must first determine the various physical conditions which actually obtain, and he must then find out which of these, if any, are essential to the production of the phenomenon. If three conditions are observed, an experiment must be made in which all of them are present, in order to make it certain that the result really follows. Then three experiments must be made with the three pairs of conditions.

Three more must then be performed with each condition present alone. Lastly, it may be necessary to make another experiment in the absence of all the conditions. In all, eight experiments may be necessary when there are three conditions. If four conditions are present, one experiment with all the conditions present, four with three conditions, six with two, four with one, and one with none, may be required—in all sixteen experiments. With one condition only, not more than two experiments are necessary, and the number is doubled for every additional condition introduced. If only ten conditions existed, more than a thousand experiments would be necessary to completely exhaust all the possibilities. Obviously, science could make little progress were such immense labour a necessity. Fortunately it is not. Past experience and natural instinct indicate to the experimenter the direction in which truth lies, and thus he is often enabled to take a short road to the end in view.

15. One great means by which labour is reduced is the employment of a suitable and probable *hypothesis*. Certain facts are known, and a hypothesis is framed regarding their explanation.

The greater the number of phenomena which a given hypothesis can explain, the greater is the likelihood of its truth. When only some facts fall in with the assumptions while others do not, some modification of the hypothesis must be made. But when new modifications have to be made for every new requirement, it is time to abandon the hypothesis and seek for another and more probable one. A good hypothesis must explain all the facts for the elucidation of which it was framed. It should also explain other known facts, and facts which subsequently become known. But, in such a case, it is customary to speak of it as a *theory*. Above all, a good theory should lead to the prediction of previously unknown facts.

It sometimes happens that different theories are each sufficient for the explanation of known facts. One theory may explain certain phenomena more easily than another can, while in the explanation of others it is more laboured. The logical consequences of the two theories must then be worked out as far as possible, and it will usually be found that at one or more points each leads to an opposite conclusion. Here experiment must step in to determine which conclusion is correct, and so to decide between the two theories. This experimental investigation is termed a *crucial test*. Very prominent examples occur in the theories of heat and light.

The tendency of scientific investigation in the present day is towards the formation of dynamical explanations of all phenomena—

towards the production of theories in which all purely physical phenomena are explained in terms of matter, and the energy which is associated with it.

Mathematical theories form an important class in which the mathematical consequences of the fundamental assumptions are rigidly worked out. When the postulates are merely expressions of known facts, the consequences of such theories may be regarded as strictly true; but in making such a statement, we must remember that all our knowledge is only approximate, being limited by the imperfections of our senses and our instruments, so that the above expression, 'strictly true,' means merely that we cannot detect deviations from the truth. The theory of gravitation is of this kind, and it furnishes us with one of the finest examples of prediction. From irregularities in the motion of the planet Uranus, Adams and Leverrier were led to foretell the existence and indicate the position of the previously unknown planet Neptune.

In other mathematical theories there is merely a partial experimental basis; for example, the dynamical theory of heat or the undulatory theory of light. Again, it is possible to work out mathematical theories of phenomena in which we know that something moves; but we may not know *what* is moving or *how* the motion is propagated. The theories of heat-conduction and of electro-dynamics are prominent examples.

As knowledge advances some theories will be shown to be false, while others will approximate more and more to certainties as the evidence of their truth accumulates.

A very important scientific method, which is in essence hypothetical, is known as the *argument from analogy*. When we perceive resemblances between different physical systems or processes, we say that they are analogous; and when any new fact is discovered regarding one of the systems, we are led by analogy to look for something similar in the others. The principle is of extreme importance in experimental work, as it indicates a promising direction for research, and so prevents aimless and often fruitless labour; and, further, the failure of an analogy may be as instructive as its success. There are many analogies between the phenomena of sound and of light; but there is nothing in sound which corresponds to polarisation in light, and the distinction is of fundamental importance.

Another extremely important aid to research is derived from the condition for stable equilibrium. This condition may be expressed as follows: A system is in stable equilibrium, under given physical conditions, when any small variation of one or more of these produces other variations which would themselves, as causes, produce changes opposite to the first.

[It is easy to see that this statement really expresses the condition for stable equilibrium. For, if one variation produced another variation which caused further variation of the first kind, this additional variation would cause more variation of the second kind, and so on reciprocally. Therefore the variation, once started, would constantly increase, or, in other words, the presumed state of equilibrium is unstable.]

The equilibrium of a body supported by the hand affords a ready illustration. Increased pressure of the hand upon the body causes an upward motion of the body. Conversely, the independent communication of upward motion to the body diminishes the normal pressure between the hand and it.

Another example is furnished by water, which is physically stable (under ordinary circumstances) below its maximum-density point, and which, at temperatures below its maximum-density point, contracts when the temperature is raised. Hence, in accordance with the above principle, we can assert that sudden diminution of volume caused by the application of pressure will produce a fall in temperature. Again, sudden elongation raises the temperature of indiarubber; therefore the increase of temperature of the stretched indiarubber makes it shrink. Lord Kelvin has proved experimentally that both these results are true. We shall see subsequently that they follow as consequences of the dynamical theory of heat.

16. In all observations, alike of natural processes and of experimental results, errors of observation are almost certain to arise. Such inaccuracies are as likely to be in excess as in defect, and are much more likely to be small than to be large, while a very large error will practically never occur at all unless the method of observation is an extremely objectionable one. To get rid of these errors we must make a sufficient number of independent observations. In any one observation we do not expect the result to be correct; but there is a certain numerical quantity, called the *probable error*, such that the actual error is as likely to be greater than it as to be smaller than it. If each observation is made under conditions precisely similar to those of another, the probable error of each is the same. In this case we simply take the arithmetical mean of all the observations, and this gives the result which is most likely to be near the true value. But if each observation is not made under precisely similar circumstances, the probable error of each will in general be different, and its value will be known for

each from the known experimental conditions. The most probable value is now found by a mathematical process known as the method of least squares.

In addition to errors of observation, there may be errors which tend always in one direction, so that the result obtained is either too large or too small. Such errors are due usually to the instrument or to the method of observation used, and are generally termed *instrumental errors*. Under this heading may be included errors due to peculiarities of the observer, and the correction to be applied is termed the *personal equation*. When only differences of values of a quantity under different circumstances are required, such errors frequently affect each observation alike, and so may be neglected. But, in general, they must be eliminated by varying the instrument and the observer, and combining the results so as to get the most probable value.

17. Most frequently in physical inquiries we have to investigate the variations of some quantity consequent upon the variation of another. The experiment may often be so arranged as to give a continuous record of the mutual variation of the two, as in the case of the self-registering thermometer and similar instruments; or even a simultaneous and continuous record, as in the case of the rise of water in the wedge-shaped space between two vertical glass plates (§ 98). But, more generally, the results of a few separate experiments are given, each of which records one definite value of the one quantity corresponding to one definite value of the other. From these detached results the law connecting the variations of the two, or, rather, an approximate law must be found, the approximation to be so exact that the result given by the law for intermediate values of the quantities shall not differ from that which may be determined afterwards by experiment by an amount greater than the possible error of observation. Such a relation is termed an *empirical law*, and the formula expressing it is called an empirical formula.

When the observed values are sufficiently close together, such formulæ enable us to find intermediate values with considerable accuracy, and also to find values altogether outside the experimental range for a short distance; but, if pushed too far, the formulæ will give values differing more and more from the truth.

Instead of seeking for an empirical formula, we might plot a curve, the abscissæ of which represent the values of the quantity to which we give arbitrary values, while the ordinates represent the values of the quantity whose variation we observe. The points so obtained will not generally lie on a smooth curve because of observational errors, but a curve drawn freely through them, leaving on the whole as many points on one side of it as there are on the other, will be fairly free from such inaccuracies, and will generally give a better approximation to the truth than the actual points themselves give. This is known as the *graphical method*, and is largely used by experimentalists. For example, the square of the time of oscillation of a simple pendulum is proportional to the length of the pendulum. Hence, if we make the abscissæ represent lengths

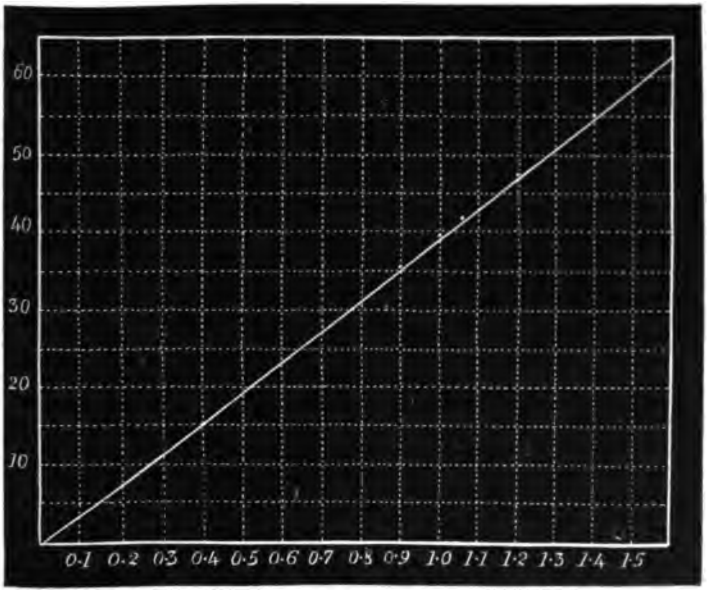


FIG. 1.

while the ordinates represent the squares of the periods of oscillation, we should get a straight line passing through the origin. In Fig. 1, which is drawn from the results of actual experiment, it will be seen that the points, while they all agree very well with each other, do not give the proper ratio of the quantities. The straight line is inclined at the proper angle. From the close agreement of the different results, we infer that there is little error of observation, but the discordant inclination of the line shows that there must be an instrumental source of error. If a series of experiments were made at parts of the earth's surface where gravity had sensibly different values, we should get a series of straight lines all passing through the origin, but all inclined at different angles. All such series of curves may be regarded as contours of a surface, and the subject is of such importance in physics as to merit discussion in a separate chapter.

Very often the form of the curve obtained by the graphical method indicates at once a suitable empirical formula. When such a formula has been obtained, by whatever process, the next duty of the investigator is to frame a hypothesis which shall sufficiently explain it.

CHAPTER III.*

THE GRAPHICAL METHOD IN PHYSICS.

18. We say that a line has only one dimension, for only one number, with the proper sign attached, is required to completely specify the relative position of two points on the line. A surface has two dimensions, for two directed lengths define the position of a point on it with reference to any other point taken as origin. Thus we speak of one point on the surface of the earth as being so much north or south, and so much east or west of another. Three directed lengths determine the relative position of two points in space, that is, in extension of the third order. Thus we speak of the length, breadth, and thickness of a solid.

It must be observed that three *lengths* are not *necessary*. *One* of the given conditions must be a length, but the others may be angles. For example, instead of saying that one mountain-top is so far north or south, so far east or west, and so much higher or lower, than another, we might give the distance between the two peaks and their relative altitude and azimuth.

The intersection of any surface, which has a constant characteristic, with the surface of a solid is called a *contour-line*. A good illustration is furnished by the contour-lines in an Ordnance Survey map. Any such line is the intersection of a surface, all points of which are at a uniform height above sea-level, with the surface of the solid earth. An Ordnance Survey map gives a very good idea of the distribution of places on the earth's surface as regards height, as well as with reference to latitude and longitude, and the information is more and more minute as the number of lines is greater. In other words, contours enable us to represent on a plane surface the mutual relations of *three* quantities—which is the essence of the graphical method.

This gives the key to the great importance of the theory of contours in physical science. For, if the physical condition of a substance is completely defined when the simultaneous values of three of its properties are given, we can construct a solid the surface of which represents all possible conditions of the substance—just as we can construct a model of the earth's surface in terms of latitude, longitude, and height above sea-level.

---

* This chapter may be omitted in a first reading.

2

19. The contour lines with which we are most familiar in nature are those formed by the intersection of level surfaces with the surface of the earth. The line of sea-board is one such contour line. The numbers marked upon maps or charts which give the height above, or depth below, sea-level indicate contour-points. When the points are taken sufficiently close together, and continuous curves are drawn through points of uniform height, we get contours, as in the Ordnance Survey maps. Such contours coincide very closely with the contour-lines formed by level surfaces. They do not exactly coincide, because of the non-spherical shape of the earth, and because of its rotation, etc. But the assumption that contour-lines of constant *level* are lines of constant height over sea-level will not introduce appreciable error, so long as the area on which they are drawn is small in comparison with the whole surface of the earth. The kinetic energy which is acquired by a body in falling freely from any point on one level line to any point on another level line is constant.

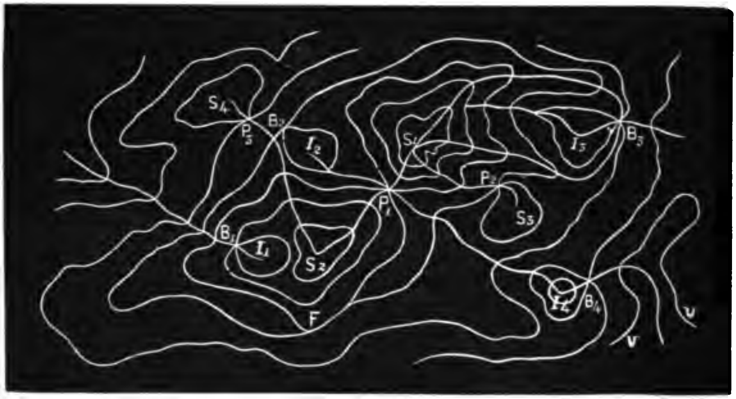


FIG. 2.

Suppose the earth to be entirely submerged underneath the surface of water, so that we have only one region, and that a region of depression below the surface of the water, to consider. If we suppose further that the water is slowly absorbed by the solid matter of the earth, regions of elevation will be formed gradually, until finally we shall have again only one region, and that a region of elevation. Before a region of elevation is formed, we have a *summit* appearing above the water-level; and when the water subsides out of a region of depression, we have a lowest-point or *imit* appearing.

The number of regions of elevation and depression may vary in two ways. Two regions of elevation may run into each other as the water sinks. The point where they first meet is termed a *pass* or *col* (see Fig. 2; P_1, P_2, etc.). Again, a region of elevation may throw out arms which run into each other, and so cut off a region of depression. The point where they first meet is termed a *bar* (B_1, B_2, etc.). The contour-line for a level immediately underneath that corresponding to the bar has a closed branch within the region of depression cut off. Thus the closed curve at I_4 is part of the contour-line UV. In the map of such a country, a pass occurs at the node of a figure-of-eight curve, or *out-loop* curve (as Professor Cayley has termed it), while a bar occurs at the node of an *in-loop* curve. If, in the diagram, the Ps represented bars and the Bs represented passes, the map would be that of an inland basin; so that, in the map of such a country, a pass is represented by the node of an *in-loop* curve, and a bar corresponds to the node of an *out-loop* curve. If there were any advantage in having passes and bars always indicated by the node of the same kind of curve respectively, this could be attained by affixing the positive sign—not constantly to the region on the same side of the level surface but—to the region towards which (or from which) the surface is moving at any instant.

As a particular case, two regions of elevation may run into each other at a number of points simultaneously. Of these points, one must be taken as a pass and the others as bars. Singular points may also occur, when, for example, three or more regions of elevation meet. Such points are called *double*, *treble*, etc., passes. Multiple bars may similarly occur.

Before a pass can be formed there must be two summits, and for every additional pass there is another summit. Thus the number of summits is one more than the number of passes. So also the number of imits is one more than the number of bars.

Slope-lines are lines drawn at right angles to the contour-lines; and, evidently, the steepness of a district is indicated on a map by the closeness of the contours. Two kinds of slope-lines are of special importance. These are the slope-lines drawn from summits to passes or bars, and from passes or bars to imits. The former can never reach an imit, and are termed *water-sheds.* The latter can never reach a summit, and are called *water-courses.*

A perpendicular precipice is indicated on a chart by the running together of two or more adjacent contour-lines (*F*). An overhanging precipice is indicated by the lapping of the upper-level line over a lower-level line.

20. Since we can represent the physical state of a substance with regard to three quantities by means of a surface, it follows that we can deduce from the contours of the surface the nature of the variation of the properties of the substance—the methods being identical with those of the preceding paragraph. We shall consider, as a particular example, the surface which represents the state of water-substance with regard to volume, pressure, and temperature. This surface was first studied by Professor James Thomson.

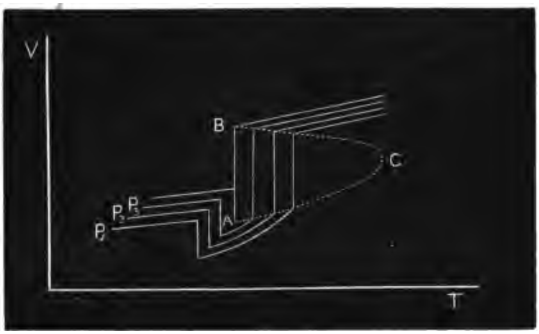


FIG. 3.

Suppose the surface to be cut by a plane of constant pressure, say P_1. We thus get a contour-line, the general nature of which is indicated in Fig. 3. At a low temperature the volume is small, the substance being in the solid state. As the temperature rises, the substance expands until liquefaction occurs. The volume then diminishes without rise of temperature, until the substance is completely liquefied. The temperature then rises while the volume diminishes, until the maximum-density point is reached. After this, expansion accompanies rise of temperature up to the boiling-point. At this stage the volume rapidly increases, while the temperature remains steady until the substance is entirely in the gaseous state. Beyond this point both increase together. The contours for slightly less pressures (P_2, P_3) are approximately parallel to P_1, but lie entirely above it; the reason being that at a given temperature the volume increases as the pressure diminishes, while the freezing-point is lowered, and the boiling-point is raised, by pressure. The freezing and boiling points approach, and finally coincide as the pressure diminishes. At lower pressures the substance changes directly from the solid into the gaseous state. The line AB in the figure above indicates the *triple-point* temperature, that is, the temperature at which portions of the substance in the three states—solid, liquid, and gaseous—can exist together in equilibrium. The increase of volume on vaporisation continually diminishes as the pressure is raised, until finally (at C) the process of vaporisation ceases. The temperature at which this occurs is called the *critical temperature.* There may also be a critical temperature for the solid-liquid condition. That is to say, there may be a temperature below which no amount of pressure will lower the freezing-point sufficiently to admit of liquefaction.

21. As another very appropriate example, we may consider the first diagram given in the chapter on electric currents. In that case the analogy to ordinary contour lines is very close. A point on a contour map represents position on the surface of the earth; a point on the diagram represents position on a conducting sheet. The lines of constant potential on the diagram correspond to the lines of constant height on the map. The lines on the diagram which cut the equipotential lines at right angles correspond to stream lines on the map. They pass from a point which corresponds to a summit to a point (the centre of the circle) which corresponds to an imit, and, by their closeness together, they indicate the density of the current. Further, as the closeness of contour lines to each other indicates parts of the earth's surface at which water would flow rapidly downwards, so also the closeness of equipotential lines indicates parts of the conducting sheet at which the current-density is strong.

Again, if a series of observations be made on the refractive indices of various aqueous solutions of a salt for light of several wave-lengths, the results may be graphically represented by points, as in Fig. 4. The curves shown in the figure, passing on the whole through these points, are contours of a surface which might be constructed by erecting perpendiculars to the plane of the diagram of lengths representing the strengths of the solutions. By the aid of the diagram we may estimate with fair accuracy the refractive index of a solution whose strength differs somewhat from those which were actually observed.

Lastly, we may refer to the figure illustrating dichroism (Chap. XV.). In it are given two contours of three different surfaces representing the relation between intensity of transmitted light and thickness in the cases of three substances of different absorptive powers.

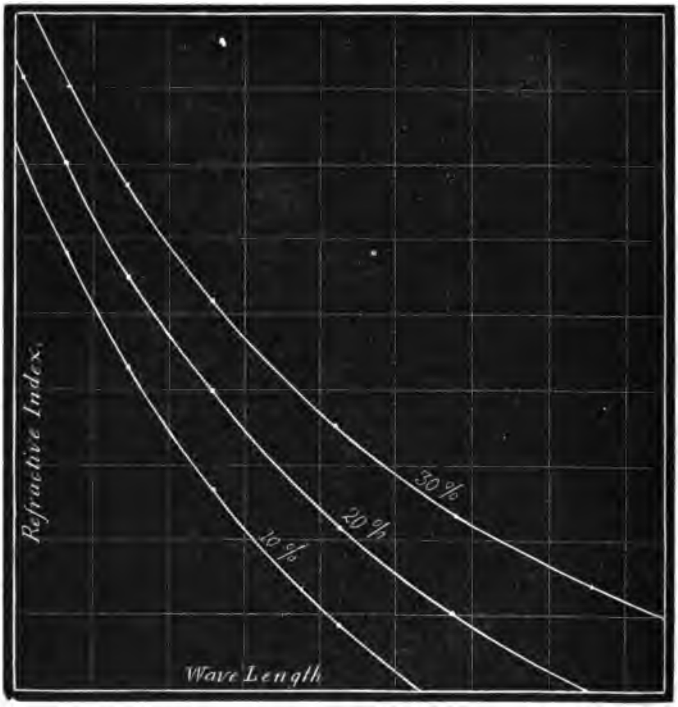


FIG. 4.

Thus small diagrams like those above considered communicate, at a glance of the eye, information which otherwise might require pages of printed matter for its adequate treatment.

CHAPTER IV.

MOTION.

22. In the preceding chapters we have had occasionally to consider questions connected with moving matter. We must now deal with this subject systematically; but, before doing so, it is desirable to discuss first the science which, dealing with motion in itself, altogether apart from any question as to what is moving or why it moves, is usually termed *Kinematics*. The idea of *time* enters into it, in addition to the idea of space, with which alone geometry is concerned.

Though every quantity, whatever be its nature, has magnitude, no quantity can be said to be large or small *absolutely*. When we speak of the size of any body we mean its size relatively to the size of some other body with which we compare it. A yard is large if we compare it with an inch; it is small when compared with a mile. In the former case the number which represents it is more than 60,000 times larger than the number by which it is represented in the latter case. A mere number is therefore useless as regards the statement of magnitude; it must be accompanied by a clear indication of what the. thing measured is compared with. The quantity in terms of which the comparison is made is called the *unit*, and the number which tells how often this unit is contained in a given quantity is called the *numeric*.

All kinematical quantities may be made to depend upon *two* units only. These are the units of *length* and *time*. Thus speed (§ 25), being measured by the distance traversed in a certain time, depends upon the unit of length directly, and upon the unit of time inversely. Hence by doubling the unit of length we double the speed unit, and therefore halve the numeric of any given speed; whereas by doubling the unit of time we halve the speed unit, and therefore double the numeric of a given speed. Again, acceleration (§ 26), being measured by the increase of speed in a certain time, depends upon the unit of speed directly and upon the unit of time inversely;

that is, it depends directly upon the unit of length and inversely upon the square of the unit of time. The manner in which the fundamental units are involved in any quantity determines the *dimensions* of that quantity. If L and T represent the units of length and time respectively, the dimensions of speed and acceleration are indicated by the symbols $[LT^{-1}]$ and $[LT^{-2}]$ respectively.

23. *Position.*—The *position* of a point in space is completely determined when *three* independent conditions are given, each of which it satisfies. And its position can only be given *relatively* to that of another point, for we do not know any point of which we can assert that it is absolutely fixed. We may say that one point is so much to the north or south of another, so far to the east or west of it, and so much higher or lower than it; or we may say that it is so far distant from the other, that the line joining the two is inclined at a certain angle to the vertical, and that the vertical plane through the two has a given inclination to the vertical north-and-south plane.

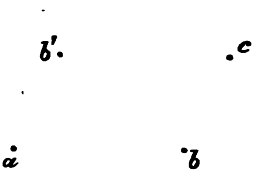


FIG. 5.

24. *Displacement.*—When a point moves from one position to another, we speak of its displacement from its first position. Two ideas are essentially involved in the term—the idea of *magnitude* and the idea of *direction*. If we know only that one point is distant three feet from another, we cannot tell what position it occupies on the spherical surface all points of which satisfy this condition. Other two conditions are required to fix the direction also, that is, to determine displacement; and no other conditions are necessary, for we do not inquire what time was occupied in the passage from the one position to the other.

Addition of two displacements is effected when we find the single displacement, which produces the same result as the two do when applied consecutively. The displacement from a to b may be ted by $\overline{ab}$, and that from b to c by $\overline{bc}$. Then $\overline{ab}+\overline{bc}=\overline{ac}$. he displacement denoted by $\overline{bc}$ might have been performed first. The effect would be to transfer a to b', which is a point such that $\overline{ab'}$ is equal and parallel to $\overline{bc}$. This follows since a displacement does not involve the idea of position, but only the ideas of magnitude and direction; in fact, ab' is the *same* displacement as $\overline{bc}$. And, similarly, $\overline{b'c}$ is the same as $\overline{ab}$; so that $\overline{ab'}+\overline{b'c}=\overline{bc}+\overline{ab}$ $=\overline{ac}=\overline{ab}+\overline{bc}$. Also, since a displacement is reversed when its direction is reversed, we have $\overline{ab}=-\overline{ba}$; and the ordinary laws of algebra apply to addition and subtraction of displacements. The *lines* ab, bc, etc., may be used to represent the *displacements* $\overline{ab}$, $\overline{bc}$, etc.; for a line involves necessarily, and only, magnitude and direction.

A line, given in magnitude and direction, is called a *vector;* and any quantity which, like a displacement, requires for its complete representation a directed line is called a *vector quantity.* In the course of this chapter, and of Chap. IV., we shall get various examples of such quantities. A quantity which is independent of direction—which merely possesses magnitude—is called a *scalar* quantity; that is, it is completely determined by measurement on a suitable scale. Sometimes a moving point may lie on a given surface or a given curve, and it is then frequently convenient to speak of its displacement *on the surface* or *along the curve.* This means that the magnitude of the displacement is to be measured *along* the surface or curve.

25. *Speed and Velocity.*—The displacement of a point may be constant, or it may vary. If it varies, we say that the point is in motion: so that by *motion* we mean change of relative position. The motion of a point is essentially a *translation,* for it has no separate parts which can rotate relatively to each other. Its position, we have seen, is not fixed unless three independent conditions are given. The removal of one restraining condition leaves the point more free to move than before; and a point, the motion of which is unrestrained, is said to have three degrees of translational freedom.

We say also that the rate of motion is constant when the point passes over equal portions of its path in equal times. The mere *magnitude* of this rate is called the *speed* of the moving point; but, when the direction is considered also, the term *velocity* is used. Thus we speak of a speed of four miles per hour and of a velocity of four miles per hour in a northerly direction. When the speed is constant, the distance traversed is proportional to the time during which the motion has continued; and, in a given time, with different constant speeds, the distances traversed are proportional to the speeds. Hence, if s be the distance, v the speed, and t the time, we find that s is proportional to the product of v and t. We may therefore write

$$s = vt \tag{1}$$

if we choose to regard unit speed as that speed with which unit distance is described in unit time.

If a point passes over a certain distance in a certain time with varying speed, it is always possible to find a constant speed with which the same distance might have been described in the same time. This is called the *Average Speed* of the point. The above equation obviously applies to the case of average speed.

26. *Acceleration of Speed.*—When the velocity of a moving point varies it is said to be *accelerated*, and the time-rate of its change is called the *acceleration.* Meantime we shall suppose that the change affects the magnitude only and not the direction: that is, we shall investigate varying motion in a straight line. We shall also limit ourselves to the consideration of constant acceleration. In this case the increase of speed is strictly proportional to the time during which the acceleration has acted; and, in a given time, with different constant accelerations, the increase of speed is proportional to the acceleration. Hence, if a be the acceleration, while v and t represent the speed acquired from rest and the time, we see that v is proportional to the product of a and t. We may therefore write

$$v = at$$

if we choose to regard unit acceleration as that acceleration under which the speed is increased by unity in unit time. If the point were originally moving with speed V we would have

$$v = \mathrm{V} + at. \tag{2}$$

Since the acceleration is constant, the average speed is one-half of the sum of the initial and final speeds. Therefore, by equation (1)

$$s = \tfrac{1}{2}\{\mathrm{V} + (\mathrm{V} + at)\}t = \mathrm{V}t + \tfrac{1}{2}at^2. \tag{3}$$

From (2) and (3) we get $v^2 = \mathrm{V}^2 + 2\mathrm{V}at + a^2t^2 = \mathrm{V}^2 + 2a(\mathrm{V}t + \frac{1}{2}at^2)$; that is,

$$v^2 = \mathrm{V}^2 + 2as. \tag{4}$$

27. *Motion in a Vertical Line under Gravity.*—The above results are sufficient for the solution of all problems regarding motion, under constant acceleration, in a straight line. As an important particular case, we may take the motion of a body in a vertical line under the action of gravity. If g represents the acceleration due to gravity (roughly 32·2 feet, or 981 centimetres, per second per second), while s is measured *upwards* from the surface of the earth, we must

put $a = -g$ in the above equations. We then find from (2) the value of the speed at a given time when the initial speed is known, the value of the time at which a given speed is attained, or the initial speed requisite for the attainment of a given speed at a given time; from (3) we get the height reached at a given time, and so on. The two values of t in (3) [and of v in (4)] correspond to upward and downward motion respectively. By putting $v=0$ in equations (2) and (4) we find that the time taken to rise to the greatest height is

$$t=\frac{\mathrm{V}}{g},$$

and that the value of this height is

$$s=\frac{\mathrm{V}^2}{2g}.$$

By putting $s=0$ in (3) we see that the body falls to the ground after an interval of time

$$t=\frac{2\mathrm{V}}{g}.$$

Therefore the body takes as long to fall from rest through a given distance as it took to rise through that distance to the resting-point. More generally, it takes equal times to pass up or down over a given part of its path. Of course, in this statement, the resistance of the air is neglected. In the case of small bodies this resistance would largely affect the motion.

If, in (3), we put $\mathrm{V}=0$, we see that the acceleration varies in inverse proportion to the square of the time when s is constant; and thus we verify the remark made in § 22 regarding the dimensions of acceleration, and justify the double use of the phrase 'per second' made at the commencement of the present section. This point cannot be too carefully attended to when changing from one set of units to another: thus the value of the acceleration due to gravity in feet per hour per hour is 32×60^4, *not* 32×60^2.

28. *Resolution and Composition of Velocities and Accelerations.*—Velocities and accelerations, since they are vector quantities, are to be compounded (*i.e.*, added; see § 24) and resolved into components in accordance with the laws of composition and resolution of these quantities. Hence, if a point is subjected to a series of simultaneous velocities which are represented by all the sides AB, BC, etc. (Fig. 6), of a closed polygon, taken in the same order round, except one, the resultant velocity is represented by the remaining side taken in the opposite direction round. This theorem is known as the 'polygon of velocities.' It follows that, if the various velocities

to which a point is subjected are representable by all the sides of a closed polygon taken in order, the point is at rest. For the resultant of all but one is equal and opposite to that one.

In the particular case of two velocities $\overline{AB}$ and $\overline{BC}$, the resultant is $\overline{AC}$—the third side taken in the opposite direction round. This theorem is known as the 'triangle of velocities.' But, since the vector $\overline{AD}$ is identical with the vector $\overline{BC}$ (Fig. 7), we may say that the resultant of two velocities represented by adjacent sides of a parallelogram is the diagonal drawn from the same point. In this form the theorem is known as the 'parallelogram of velocities.'

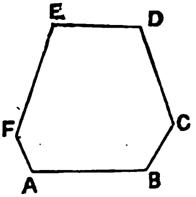


FIG. 6.

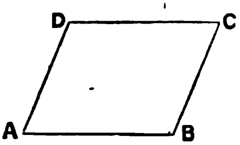


FIG. 7.

In order to resolve a velocity into any number of components we have merely to reverse the above process.

As a particular case, if we have to find the resolved part of a velocity $\overline{AC}$ (Fig. 8) in a direction AB making an angle θ with AC, we require to know first the direction of the other component. It is usually understood that the other component is to be at right angles to the first, in which case we have $AB = AC \cos \theta$, and $BC = AC \sin \theta$; that is, either of the short sides of a right-angled triangle represents the component, in its own direction, of the quantity which is represented by the hypothenuse. Consequently the square of the total speed is equal to the sum of the squares of its rectangular components.

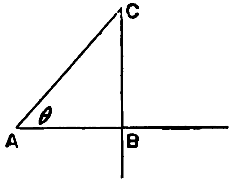


FIG. 8.

29. *Curvature. Acceleration of Velocity.*—In sections 26 and 27 we have supposed the direction of motion to be unaltered. When the direction changes, the path of the moving point is said to be 'curved.' The tangent to the curve gives the direction of motion at any instant, and the limiting value of the ratio of the angle between two tangents at near points to the length of path between these points as they are taken closer and closer together and finally coincide is called the *Curvature* of the path at that place. Thus the curvature is the rate of change of the direction of motion per unit length of the curve.

To obtain a measure of the angle between two lines in a plane (and we are here limiting ourselves to the case of motion in a plane) draw a circle (Fig. 9), of any radius r, from the point of intersection of the lines as centre. The angle θ is measured by the ratio which the length, s, of the arc of the circle intercepted between the lines bears to the radius. It follows that the ratio of θ to s is equal to $1/r$,

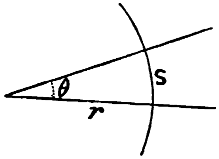


FIG. 9.

and is therefore constant for a given circle no matter how large or how small θ and s may be. Hence the curvature of a circle is the reciprocal of the radius.

Now, it is always possible to draw a circle the curvature of which is the same as that of a given curve at a given point. This circle is called the *circle of curvature* at that point; its radius is called the *radius of curvature;* and the reciprocal of its radius measures the curvature of the given curve at that point.

In considering change of velocity as dependent on change of direction, it will be convenient to assume at first that the speed is constant, and also that the rate of change of direction is constant. In other words, we shall investigate the case of uniform motion in a circle.

Since the speed is constant, there can be no component of acceleration in the direction of motion; hence the change in the direction of motion must be due to an acceleration whose direction is at right angles to the direction of motion. Let BCDE (Fig. 10) represent the path of radius r, and let BC be a portion of it so small as to be practically straight. When the moving point is at B the direction of motion is along BP, and the acceleration must be directed towards the centre O, since the point moves to that side of BP on which O lies. Because the initial motion has no resolved part to or from O,

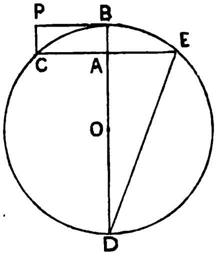


FIG. 10.

the distance PC (or BA) by which the point moves towards O under the central acceleration must be described as if there were no initial motion. Hence, if a denote the acceleration, we get—by (3), § 26—$BA=\frac{1}{2}at^2$ where t is the time of description of the small arc BC; and we have also—by (1), § 26—$BP=AC=vt$ if v be the uniform speed in the circle. Now, from geometry, $AC^2=AB.AD$, and AD is practically equal to $2r$. Therefore

$$a=\frac{v^2}{r}.$$

30. *Motion of Projectiles.*—The velocity of a projectile in a vertical plane may be resolved into its horizontal and vertical components. There is no horizontal acceleration, and so the horizontal speed is

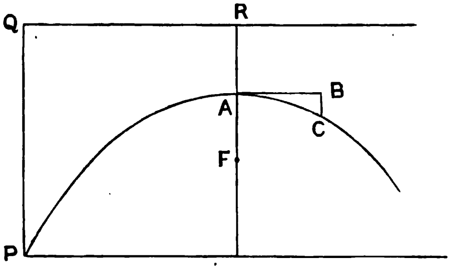


FIG. 11.

constant. There is constant downward acceleration, and so the vertical component of the velocity is in accordance with the results given in §§ 26, 27. Thus, if v_1 and v_2 represent respectively the horizontal and vertical components of the speed of projection, the quantity $v_1 t$ is the horizontal distance of the projectile from the point of projection at the time t reckoned from the instant of projection, and the quantity $v_2 t - \frac{1}{2}gt^2$ is the corresponding vertical distance. The greatest height reached is $v_2^2/2g$, and the total time of flight is $2v_2/g$. The horizontal range, being the product of the constant horizontal speed into the time of flight, is $2v_1 v_2/g$.

The nature of the path is, of course, totally unaffected by the position of the point of projection, and so we may assume that the particle is projected horizontally from A with speed v_1, along AB. If, with no vertical acceleration, the point B would be reached in time t, we have $AB = v_1 t$. But in this time the point has fallen through a vertical distance $BC = \frac{1}{2}gt^2$. Therefore

$$AB^2 = 2\frac{v_1^2}{g}BC;$$

that is, the square of the horizontal distance traversed is proportional to the vertical distance. This is a well-known property of a parabola with its vertex at A; its axis vertical; its focus, F, at a distance $v_1^2/2g$ below A; and its horizontal directrix, QR, at the same distance above A. Now, if v_2 be the vertical speed at P, we have seen that AM is equal to $v_2^2/2g$. Therefore $PQ = AM + AR = (v_1^2 + v_2^2)/2g = V^2/2g$ (§ 27) if V be the total speed at P; and, by a property of the parabola, this is also the value of PF, the focal distance. Hence the magnitude of PF depends only on the speed of projection, and not at all on the direction of projection. Thus, with a given speed of projection, the focus lies on a circle, the centre of which is the point of projection, and the radius of which is the distance of that point from the directrix. Since the point of projection may be anywhere on the path, this construction holds for all points. Hence, if two points on the path and the speed at one of them be given, so that the construction can be carried out for both, the intersections of the two corresponding circles will give the foci of possible paths passing through both points. Since there are in general two such points of intersection, we see that there are in general two possible paths by which, with a given speed of projection from one point, another given point can be struck. In particular, when the two circles just touch, there is only one path from the first point to the second; and no other point which is at a greater distance from the first in the same direction as the second can be reached with the given initial speed, for there is no real point of intersection of the corresponding circles.

31. *The Hodograph.*—If, from any point as origin, a line be drawn to represent the volocity of a moving point, the free extremity

28

to

clo

sul

is

th

v

tl

p

tl

h

v

[illegible] is called the hodograp
of the line [illegible] If OQ, OP (Fig. 12) represent the in
path of the moving point, [illegible] QP represents the total change of velocity.
[illegible] points of the line QP do not in g
on the hodograph, we thus see that any chord of the h
represents the total change of velocity which takes pla
interval of time between the passages through the ex
the chord, of the moving point which describes the h
Let the curve PP′ (Fig. 13) represent the hodograph in
of motion. When the points P, P′ are very close togeth
practically coincides with the tangent. Hence the dire
tangent in the hodograph gives the direction of accele
path, and the rate of motion in the hodograph gives th

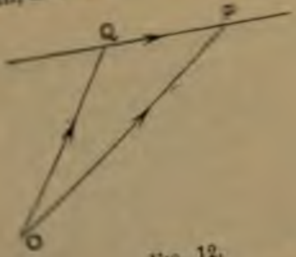


Fig. 12.

value of the acceleration in the path. As a simple
value of this method of representation we may co
case of uniform motion in a circle (§ 29). Since t
stant, the hodograph is also a circle; and, since
motion in the path revolves uniformly, the hodog
formly described—the angular rate of descriptio
for both circles. Hence the ratio of the circumf
the ratio of the speeds. If r be the radius of t
the speed in the path, and therefore represents
the speed in the path, this condition gives $r/v=v/a$, where a
hodograph, this condition gives $r/v=v/a$, where a
hodograph. This gives, as before, $a=v^2/r$; and
acceleration is towards the centre of the path,
motion in the hodograph is parallel to the inwar
the path.

The [illegible]graph in the case of uniform mo
line [illegible], for the end of the line drawn
[illegible] velocity is fixed. And in the
[illegible] on in a straight line, the hod
[illegible] ugh the origin, described w

Combining these two cases, we see that, in the case of motion of a projectile, the hodograph is a vertical straight line, described with uniform downward speed and displaced from the origin through a horizontal distance which represents the constant horizontal speed of the projectile.

32. *Moments.*—The moment of any quantity is the measure of its importance with regard to the production of some effect. The moment of any directed quantity (which may be indicated by the line AB) with reference to revolution about a point O is proportional to its own magnitude and to the distance of O from its line of action. If we define unit moment as the moment of a directed quantity of unit magnitude about a point at unit distance from its line of action, pa is the magnitude of the moment of a quantity containing a units about a point distant p units of length from its line of action. Thus the moment of AB about O is twice the area of the triangle AOB.

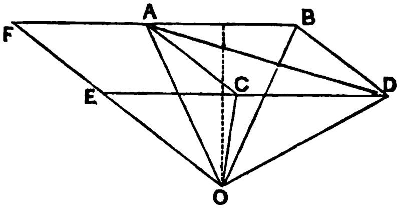


FIG. 14.

The moment of the resultant of two directed quantities is the sum of the moments of the components.—Let AC, AB (Fig. 14), be the two components; AD being the resultant. We have to prove that AOD = AOB + AOC. Draw OF parallel to CA to meet BA and DC produced in F and E. Then AOD = AOB + BOD − ABD = AOB + $\frac{1}{2}$ FEDB − $\frac{1}{2}$ ACDB = AOB + $\frac{1}{2}$ FECA = AOB + AOC.

The lines AC and AB in the figure have been so drawn that motion along them from A involves revolution in the same direction about O. Had AC been so drawn as to involve rotation about O opposite to that indicated by AB, it would have been necessary to regard one of the triangles as being *negative*. The same proof would then hold. It is usual to regard rotation opposite to that of the hands of a watch as positive.

When the direction of the quantity passes through O, its moment about O vanishes.

33. *Simple Harmonic Motion.*—When a point P revolves with

constant speed in a circle, the motion of the foot N (Fig. 15) of the perpendicular drawn from P to any fixed diameter is called simple harmonic motion. The velocity and acceleration of the point N can easily be found when the position and velocity of P is given.

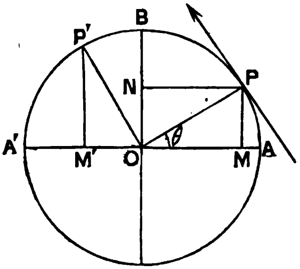


Fig. 15.

Since the radius OP is always perpendicular to the direction of the motion of P, we may use it to represent the velocity of P. And NP, since it is the component of OP perpendicular to the direction of motion of N, must then represent the velocity of N. The maximum speed is attained when N passes through O, and is equal to the speed of P in the circle.

The acceleration of N is the resolved part of the acceleration of P in the line ON. Now, the acceleration of P is (§ 29) v^2/r in the direction PO. We may therefore use PO to represent it; and, on the same scale, NO will represent the acceleration of N, its magnitude being $v^2\mathrm{ON}/r^2$—a quantity which is proportional to the displacement of N from the centre. Conversely, when the motion is rectilinear, and the acceleration is directed to a fixed point in the line of motion, and is proportional to the displacement from that point, the motion must be simple harmonic motion.

The ratio of the acceleration to the displacement is $v^2/r^2=\omega^2$, where ω is the angular speed of OP, *i.e.*, the angle traced out by OP per unit of time. But the angular speed is $2\pi/\tau$, τ being the time of a complete revolution in the circle, or of a complete oscillation in the line OB. Hence

$$\frac{\text{acceleration}}{\text{displacement}}=\frac{4\pi^2}{\tau^2}.$$

If we call B the positive extremity of the range, the angle through which OP has turned, since it coincided with OB, is called the *Phase* of the simple harmonic motion. The phase may also be measured in fractions of a whole revolution. Thus we speak of the quarter phase, etc.

The maximum distance to which N can get from O is called the *Amplitude* of the motion. It is obviously the radius of the corresponding circle.

If the motion does not commence at the positive extremity of the range, the angle through which the radius has to turn until P reaches the positive end is called the *Epoch.*

Simple harmonic motion is frequently exemplified in nature. It occurs in the vibration of stretched strings, the agitation of the luminiferous medium, etc.

34. *Composition of Simple Harmonic Motions.*—(1) *Motions in the same straight line and of equal periods.* Let OA be the line, and let uniform circular motion of P (Fig. 16) correspond to one of the given motions, of which θ is the phase. From P draw a line PQ, making an angle ϕ, equal to the phase of another of the motions, with the line OA; and let the length of PQ be equal to the amplitude of this motion. Since ϕ and θ increase at the same rate, the line OQ remains of constant length and revolves at the same rate as OP and PQ. But the foot of the perpendicular from Q on OB moves with a motion which is compounded of the two given motions; for OM and MN, where M is the foot of the perpendicular from P on OA, are the two displacements. Hence the resultant of the two motions is another simple harmonic motion, of the same period, in the same straight line. And, in particular, when the amplitudes of the two components are equal, the phase of the resultant is the mean of the phases of the components.

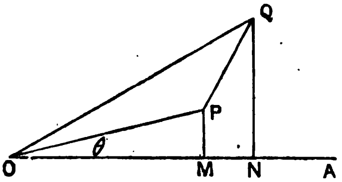


FIG. 16.

This proof is quite general, and applies to any number of such simple harmonic motions.

(2) *Two simple harmonic motions, of the same period and phase, in lines inclined at any angle.* These obviously compound into a single motion of the same period and phase. For, let OA, OB be the two inclined lines, and draw any other line OP. From P draw PM, PN, parallel to OA, OB, respectively, to meet these lines in the

points N and M respectively. If P moves along OP with simple harmonic motion, it is evident that M and N move along OB and OA with simple harmonic motions of the same period and phase. And the motion of P is the resultant of their motions.

(3) *Two simple harmonic motions in lines at right angles to each other of the same period and amplitude, and differing in phase by one quarter of a period.* A glance at the figure of last section shows that M′ moves along OA′ precisely as N moves along OB, since OP′ is perpendicular to OP. Hence the motions of N and M differ in phase by one quarter of a period. And the motion of P is their resultant. That is to say, the resultant is uniform circular motion; and the motion takes place from the positive end of the range in which the motion is one quarter of a period in advance to the positive end of the other range.

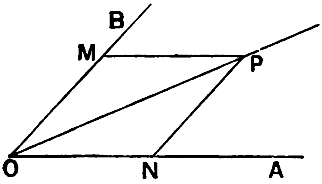


FIG. 17.

(4) *Two simple harmonic motions of equal period, and of phases differing by a quarter period, in lines inclined at any angle.* By projection of the above circle we obtain an ellipse in which conjugate diameters correspond to mutually perpendicular diameters of the circle. Hence the resultant is elliptic motion, with equal areas described in equal times (since this is so in the circle), and with a law similar to that given in (3) as regards direction of motion.

(5) *Any number of simple harmonic motions, in lines inclined at any angle to each other, and of any phase, but of equal periods.* By a reversal of the first proof of (1), we see that the line OQ may be replaced by *any* two lines OP and PQ, which revolve with the same angular velocity. Hence any simple harmonic motion may be broken up into two of the same period, which differ in phase by any given amount, and one of which has any given phase.

Let P_1, P_2, etc. (Fig. 18), be the points moving with simple harmonic motion. Let p_1, p'_1, be two points whose motions compound into that of P_1, and let their phases differ by a quarter period. Deal similarly with P_2, etc., and let the motions of p_1, p_2, etc., be all of the same phase, while those of p'_1, p'_2, etc., also agree in phase with each other, but, of course, differ in phase from the motions of p_1, p_2, etc., by a quarter of a period. Resolve all these motions into their components along two rectangular axes *ox* and *oy*. Then compound all of the same phase in each axis with each other. The result is, by (1), two simple harmonic motions in each axis differing in phase by a quarter period. Now combine each motion in *oy* with the motion in *ox* which is of the same phase, and we have ultimately two simple harmonic motions which differ in phase by a quarter period, and take place in lines which are in general inclined to each other at a finite angle. These, as we have seen, combine into elliptic motion.

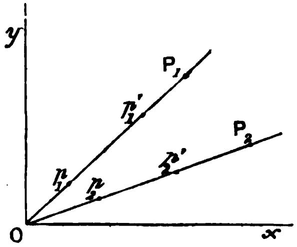


FIG. 18.

35. *Wave Motion along a Line.*—Let a point vibrate up and down the line *oy* with simple harmonic motion about the point *o*, and let the paper be drawn along behind it at a uniform rate in the direction *xo*. The point will trace out a curve (indicated in the figure) which exhibits the simplest form of wave motion. The distance of a crest from the line *ox* is the amplitude of the wave, and the distance from crest to crest measured parallel to *ox* is the *wave-length*. If other points vibrate simultaneously with different periods and different amplitudes, various series of such waves will

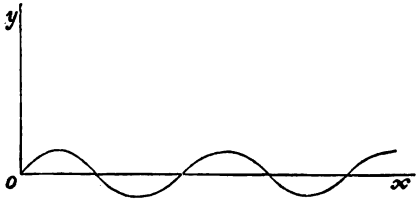


FIG. 19.

be described on the diagram as the paper is drawn along; and we may add together the displacements due to each motion, and so get the single resultant wave caused by the simultaneous occurrence of the several simple waves. In particular, suppose that two series of equal waves coincide for the moment in the form shown in the diagram, and that they move with equal speeds in opposite directions along *ox*. The points at which the resultant curve crosses the line *ox* remain fixed, for any such point is as much lowered by the backward motion of one wave as it is raised by the forward motion of the other. They are points at which the oppositely moving waves always meet in opposite phases, and are called *nodes*. Points midway between successive nodes are points at which the component waves meet in like phases, and so on. If the resultant crest occurs between two given nodes at one instant, a hollow will be found between them when the component waves have each moved through half a wave-length. Thus the resultant wave oscillates up and down between alternate nodes. It is therefore called a *stationary wave*. This result is of importance in the theory of the vibrations of stretched strings, etc. If a stretched string be fixed at one end, while waves are sent along it from the other to be reflected at the fixed end, the direct and reflected series *interfere* in the manner described.

36. *Rotation.*—While a mere *point* can have translational motion only, a rigid body (a body the parts of which cannot suffer relative displacement) is free to *rotate* also unless three points of it, which do not lie in the same straight line, are fixed. Three such points being fixed, the body is devoid of all freedom to move. If two points are fixed it can rotate about the line which joins them, and is said to have one degree of rotational freedom. If one point only is fixed the body may rotate independently about any three mutually perpendicular axes which pass through that point—it has three degrees of rotational freedom. Finally, no point being fixed, it has, in addition, three degrees of translational freedom. The greatest number of degrees of freedom which a rigid body can have is therefore six. [A non-rigid body has distortional freedom also.]

37. *Uniplanar Motion of a Rigid Body.*—By 'uniplanar motion' is meant motion parallel to one plane. The motion of a rigid plane figure is included as a special case.

Let the motion be parallel to the plane of the paper; and let AB be the position of a line in the body before the motion occurs, while A′B′ is its position at the end of the motion. Draw AA′ and BB′; bisect them, and erect perpendiculars at their points of bisection. Let these meet in O. We have OA=OA′, and OB=OB′.

Also AOB and A′OB′ are congruent triangles, the angle AOB being equal to the angle A′OB′. Thus AB might have been moved into the position A′B′ by a single rotation, about O as centre, through the angle AOA′. Hence *any displacement of a rigid body parallel to one plane may be produced by rotation about a definite axis perpendicular to that plane.*

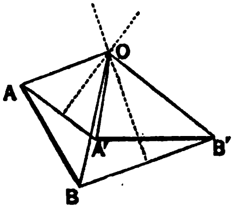


FIG. 20.

In general the body does not revolve in this way from its initial to its final position. On the contrary, each point usually describes a curve which is not the arc of a circle. In such a case we may regard the total displacement of any point as made up of a succession of indefinitely small displacements, each of which coincides with an indefinitely small arc of a circle. This circle is evidently the circle of curvature (§ 29) of the path of the moving point. Its centre is called the *instantaneous centre*, about which all points in the same plane are revolving. Thus, when a wheel rolls along the ground, the point of the wheel which is in contact with the ground is at rest for an instant—it is the instantaneous centre about which the wheel is revolving for a moment as a rigid body.

Let the point O (Fig. 21) be the instantaneous centre. The point p_1 revolving about O will come into the position p'_1. Suppose now that p_2 revolves about p'_1 as the new instantaneous centre, and that this

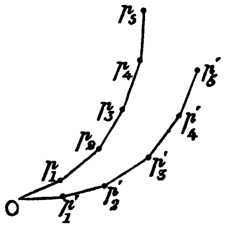


FIG. 21.

brings it into coincidence with p'_2, about which the revolution next takes place, and so on. The points p_1, p_2, etc., are points, fixed in the body, which are successively at rest, and the points p'_1, p'_2, etc., are points, fixed in space, at each of which in succession the instantaneous centre is situated. The instantaneous centre coincides for a short time with each point of both series, and passes suddenly from one to another. When the motion is continuous, the polygons Op_1p_2 . . . and $Op'_1p'_2$ . . . become continuous curves, and the instantaneous centre moves continuously along them. In the case of a rigid body, moving parallel to the plane of the paper, the line through the centre perpendicular to that plane is instantaneously at rest, and is called the *instantaneous axis;* and the curves in the figure are sections of cylindrical surfaces in the body and in space. Hence we see that *the most general uniplanar motion of a rigid body consists in the rolling of a cylinder fixed in the body upon a cylinder fixed in space.* An obvious example of this is given when a roller is drawn over the surface of the ground.

A similar statement, modified merely by the substitution of the word 'curve' for 'cylinder,' applies to the motion of a plane figure in its own plane.

Mere *translation* is a special case in which the instantaneous centre is at an infinite distance. It may be considered to consist of infinitely slow rotation about an infinitely distant axis.

38. *Motion of a Rigid Body in Space.*—First, suppose one point of the body to be fixed, and consider a sphere in the body with its centre at the fixed point. Take any two points A,B, on the surface of this sphere which occupy the positions A′,B′, respectively, at the end of the motion. The reasoning of last section applies here also, great circles of the sphere taking the place of straight lines. We thus find that the displacement might have been produced by simple rotation of the sphere about a diameter passing through the point O (Fig. 20) on the surface.

The actual motion consists in the rolling of a cone fixed in the body upon a cone fixed in space. This is at once evident if we suppose the curve Op_1p_2 . . . of last section to be drawn upon the surface of the sphere whose centre is fixed—the points p_1,p_2, etc., being successive positions of the extremity of the instantaneously fixed diameter.

Now suppose that no point is fixed. The total displacement consists in general of both translational displacement and rotational displacement; and it is clear that we may separate these, taking first one and then the other, and yet produce the same total effect. As we have just seen, the rotation leaves a set of planes (those perpendicular to the axis of rotation) parallel to their original position, and mere translation does not alter this parallelism; so that, in the final position of the body, there is one set of planes which has been unaltered as regards orientation. The required rotational displacement can be produced by revolution about *any* axis perpendicular to these unaltered planes, but the final *position* of the body will depend on the particular axis chosen. And we can so choose the axis that, after the rotation has taken place, mere translation parallel to that axis will make the body take the required position. And this can only be done in one way. For let the plane of the paper be one of the planes which are unaltered in direction, and let AO be a line in that plane, the final position of which is to be A′O′ in a parallel plane. Rotation through the angle between AO and A′O′, about *any* point in the plane of the paper will make AO *parallel* to A′O′; but AO will only be superposed upon A′O′ if the point be chosen by means of the construction of last section. And, when this superposition is effected, a translation perpendicular to the plane of the paper will bring the body into its final position. *Hence any displacement of a rigid body in space may be produced by rotation about, and translation along, a definite axis.*

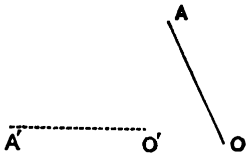


FIG. 22.

When the rotation and the translation are simultaneous, the motion is called a *twist about a screw.* Such a motion is the most general kind of motion that a body which possesses only one degree of freedom can have. The rotation and the translation do not occur *independently*, and so one degree of freedom alone is involved.

Any given motion of a rigid body in space consists of a twist about a screw, in which the axis and the linear and angular velocities are in general varying. The position of the axis at any instant is given by the line of contact of two ruled surfaces, one of which (fixed in the body) simultaneously rolls and slides upon the other, which is fixed in space.

39. *Composition of Angular Velocities.*—An angular velocity is

completely determined when its magnitude and direction are given, and these quantities may be indicated by the magnitude and direction of a line drawn parallel to the axis of rotation. Angular velocities (and accelerations) are therefore vector quantities, and their laws of composition and resolution are identical with the laws for vectors, and therefore with the laws for linear velocities and accelerations. Thus the resultant of two angular velocities or accelerations, which are represented by the two sides of a parallelogram, is represented, on the same scale, by the conterminous diagonal.

An extremely important case is that in which a body with one point fixed is revolving uniformly about an axis, and is subjected to constant angular acceleration about a perpendicular axis. In the corresponding problem regarding linear velocity and acceleration (§ 29), the magnitude of the linear velocity is unaltered while the direction of motion revolves uniformly, being always perpendicular to the direction of acceleration, which also revolves uniformly. So, in the present case, we can at once assert that the magnitude of the angular velocity will remain constant, but that its axis will revolve uniformly, and will be always perpendicular to the axis of constant acceleration. Now the axis of acceleration is always horizontal when the rotating body is a top which spins uniformly with its axis of revolution inclined to the vertical. Hence the direction of the axis of the top will rotate uniformly, and will be always perpendicular to a horizontal line (*not* a horizontal plane) which rotates uniformly. The axis must, therefore, revolve at a constant rate around the vertical.

Precession of the equinoxes is due to angular acceleration of the earth about an equatorial axis, and the peculiar motions of gyrostats have a similar explanation.

To compound two angular velocities, about parallel axes, indicated in magnitude and direction by AB and CD (Fig. 23), it is merely necessary to find a point O such that the moments of AB and CD about it are equal and opposite (§ 32). Let p_1, p_2 be the lengths

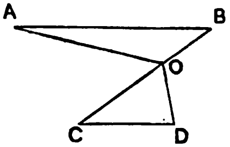


FIG. 23.

of perpendiculars from O upon AB and CD respectively. Since the angular velocity of O due to rotation about AB is numerically equal to AB, it follows that ABp_1 represents the velocity of O perpendicular to the plane of the paper; and, similarly, the velocity of O perpendicular to that plane due to rotation about CD is CDp_2; and this quantity is of the opposite sign to the former, for points moving along AB and CD revolve oppositely around O. Hence, if the areas AOB and COD are equal, the point O is at rest; and thus the locus of O—that is, a line through O parallel to AB and CD—is the resultant axis of rotation.

Similar reasoning shows that the resultant axis due to angular velocities indicated by AB and DC is a parallel line situated above AB, and that the direction of rotation coincides with that of AB, the larger of the two components.

To find the effect of the superposition of an angular velocity ω upon a linear velocity v, or of a linear velocity v upon an angular velocity ω, we may resolve the linear velocity into its two components parallel to and perpendicular to the axis of rotation. The effect of the perpendicular component v' is to shift the axis parallel to itself through a distance d, such that the speed $d\omega$ is equal and opposite to the speed v'. The parallel component simply moves the whole body in the direction to the axis, so that the resultant is a twist.

If a body is rotating simultaneously about three axes, the velocities being representable by the three sides of a triangle taken in the same direction round, the effect is that there is no rotation; but the body is translated perpendicularly to the plane of the triangle with a speed represented by twice the area of the triangle. If one of the rotations is represented by a side of the triangle taken in the opposite direction round, there is no translation; but the body rotates about an axis bisecting the two sides which were taken in the same way round. The rotation round this axis coincides in direction with the rotation about the side of the triangle parallel to it, and the angular velocity is twice as great as the velocity about the parallel side.

Corresponding results obtain in the case of angular velocities representable by the sides of closed plane polygons.

40. *Displacement of the Parts of a Non-Rigid Body.—Strain.*— A non-rigid body may alter in form, or in volume, or both. Any such definite change of shape or bulk is called a *Strain.*

Homogeneous Strain.—When all parts of a body originally similar and equal, are similarly and equally strained, the strain is said to be *homogeneous.* It follows that parallel straight lines in the unstrained body become parallel straight lines in the strained body; but, in general, the direction of the lines and the distance between them is altered. And therefore parallelograms remain parallelograms, parallelepipeds remain parallelepipeds, and any figure or surface changes into a similar figure or surface. Thus a sphere becomes an ellipsoid.

Such a strain is completely determined when we know what alteration—in magnitude and direction—has been produced in three originally non-coplanar lines. One number is required for each line to fix the change of length; and two numbers are required to determine the change of direction of each line. In all, nine numbers are in general necessary.

The simplest kind of homogeneous strain is that in which there is uniform expansion or compression in all directions. Any line in the strained figure preserves its original direction, and all lines are equally altered in length. Such strain occurs in the compression of fluids. One number completely determines it.

Next in order of simplicity is a homogeneous strain in which lines in one definite direction are unaffected by the strain; that is, the strain is confined to planes perpendicular to a definite direction, and the alteration in any one of these planes is precisely similar and equal to the alteration in any other. It is usual, therefore, to call such a distortion a *plane strain*. A circle in the unstrained figure, drawn in one of the planes of distortion, becomes an ellipse; and so a sphere in the body becomes an ellipsoid. (In the cases of equal expansion or of equal contraction in all directions in the planes of the strain, the ellipsoid is an oblate or a prolate spheroid respectively. All lines which are neither in, nor perpendicular to, these planes are altered in direction.) All diameters of the circle, which are at right angles to each other, become conjugate diameters of the ellipse; in particular, the principal axes of the ellipse were originally perpendicular diameters of the circle—from which, however, they usually differ in direction. When the plane is given, four numbers determine a plane strain of the most general kind. These might be the azimuth of one principal axis, the two principal extensions, and the rotation of a principal axis from its original position.

When the principal axes of the ellipse (called the *strain-ellipse*) into which the circle is deformed are not changed from their original directions in the unstrained body, the strain is called a *pure (or non-rotational) plane strain*. In this case the distortion consists of extension (or contraction) in two directions at right angles to each other. Any rotational, or impure, plane strain may (so far as the final effect is concerned) be produced by a pure strain superposed upon, or followed by, rotation about a definite axis perpendicular to the plane. In a pure strain every line except a principal axis has suffered rotation; and it follows from this that the superposition of two pure strains generally produces an impure strain. Hence a body may be distorted by three plane strains in succession, and yet (the strains being properly chosen) be left unstrained, but rotated through an angle about a definite axis. Three numbers completely characterise a pure strain in a given plane. These might be the azimuth of one principal axis, and the two principal extensions.

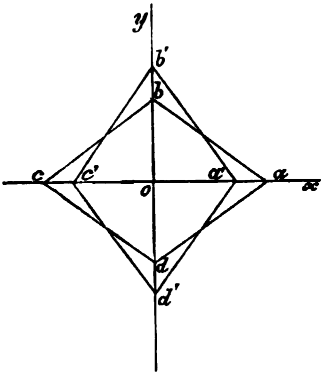


FIG. 24.

A specially important case is that in which there is no alteration of volume. This implies elongation in one direction and equal contraction in another. We may suppose that these directions are mutually perpendicular, giving a pure strain; for, as we have seen, any other strain may be assumed to consist in a pure strain followed by rotation as of a rigid body. Let oy be the direction of the elongation, and let ox be the direction of contraction. Let $abcd$ be a rhombus in the unstrained figure, which becomes the rhombus $a'b'c'd'$ in the strained state, oa being equal to ob', oa' being equal to ob, and so on. We have then ab equal to $a'b'$, with similar results

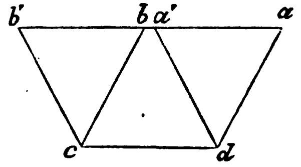


FIG. 25.

for the other sides of the rhombus. It is, therefore, obvious that there are two sets of planes in the figure which experience no alteration, except as regards position; so that (rotation excluded) the strain might have been produced (see Fig. 25) by holding fast one plane of either set—say the plane through *cd* perpendicular to the plane of the paper—and sliding all planes parallel to it through a distance proportional to their distance from the fixed one, until the originally acute angle of the rhombus becomes equal to its supplement. This motion is termed shearing motion, and the strain is called a *simple shear*.

To make the result of this shear coincide with the result of the above pure strain, we must turn the body round in the direction of the hands of a watch through an angle $b'db$, so that $b'd$ coincides with bd—that is, through an angle equal to half the difference between the obtuse and acute angles of the rhombus.

In the most general homogeneous strain, a sphere in the unstrained body becomes an ellipsoid in the strained state. Any set of three mutually perpendicular axes become mutually conjugate diameters of the ellipsoid. In particular, the three principal axes of the ellipsoid (which are called the principal axes of the strain) were originally perpendicular diameters of the sphere. These principal axes are usually rotated from their initial positions; and, as in the corresponding case of plane strain, when this turning of the principal axes does not occur, the strain is said to be *pure* or *non-rotational.* So far as the ultimate result is concerned, any impure strain may be looked upon as due to a pure strain followed by a rotation.

Any given strain, apart from rotation, may be produced by a simple shear followed by an extension (or contraction) perpendicular to the plane of the shear which in turn is succeeded by a uniform expansion (or compression). For the shear may be continued to such an extent as to give the proper ratio of the maximum and minimum axes; and the perpendicular extension will then give the proper ratio of the mean axis to each of the other two; while, lastly, the uniform expansion can be continued to such an extent as to give the proper magnitudes of the axes. As a further test of the statement we may consider the number of requisite data. For the shear, four are required—two to give the set of planes which are sheared, one to give the direction of shearing in these planes, and one to give its amount; one more is needed for the perpendicular extension; and yet another for the uniform expansion: six in all. The three additional data, which are required to make up the nine data of the most general strain, correspond to a further rotation—

two to give the axis and one to give the amount. Otherwise, the nine data might correspond to two shears and a uniform expansion.

41. *Non-Homogeneous Strain.*—So long as rupture does not occur, all displacements in a portion of matter are essentially continuous. Hence, however greatly the displacements may change throughout the body, we can always consider a portion so small that, within its limits, the strain is homogeneous.

42. *Motion of Fluids.*—While the parts of a rigid body cannot suffer relative displacement, the parts of a non-rigid solid can be displaced relatively; but the magnitude of any displacement is not unlimited, for, if it be too large, the body will be ruptured. In an infinite expanse of fluid there is no limit to the possible increase of distance between two originally near parts.

A *Line of Flow* in a moving fluid is defined as a line so drawn that its direction at any point coincides with the direction of motion of the fluid at that point. It may, or it may not, be the actual path of any particle of the fluid. In illustration of this we may consider the motion of points in a spinning top (§ 39). At any instant the line of flow of a given point is a circle drawn round the axis of the top. But the axis is itself in motion, so that the path of the point merely coincides with the line of flow for an indefinitely small distance. Similarly, when the lines of flow in a fluid are in motion, the path of any particle only coincides with one line of flow for an indefinitely small interval of time.

When the lines of flow are fixed, so that they are actual paths of particles, the motion is said to be *steady*, and the lines are called *stream-lines.*

If lines of flow are drawn through all points of a closed curve which does not lie on a surface of flow, a *Tube of Flow* is formed. None of the fluid inside such a tube ever passes out of it, and none ever enters it from the outside.

Whatever be the nature of the strain throughout a solid, we may consider, instead of the total strain, the strain produced in a given indefinitely small period of time. The displacements in that period are evidently proportional to the instantaneous velocities of the various parts. And, hence, all the results which we have obtained regarding displacements in a non-rigid solid have a direct application in the discussion of fluid motion.

Just as rotational strain may exist in a solid, so, in a fluid, there may be rotational, or, as it is termed, *Vortex Motion.* A line, drawn in the fluid so that its direction coincides at any point with the direction of the axis of rotation at that point, is called a *vortex line.* And a tube formed by vortex lines drawn through all points of an infinitely small closed curve forms a vortex tube, and is said to enclose a *vortex filament*. In most cases of fluid motion in which vortices exist, the vortically moving parts occupy only a small proportion of the whole volume of the fluid.

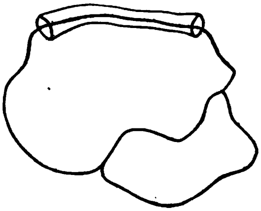


FIG. 26.

If s represents an infinitesimal length of a curve drawn in a moving liquid, while v represents the velocity parallel to the curve at the point where s is measured, the sum of the quantities vs for all points of the curve is called the *Circulation* along the curve. If we surround the curve by a tube (Fig. 26), the sectional area of which is numerically equal to the speed along the curve, the volume of that tube is numerically equal to the circulation along the curve. The positive or negative sign must be attached according as the circulation is in the positive or negative direction round the curve. When the curve is closed and lies in a plane, we may speak of the circulation 'round the enclosed area.' It is evident that the circulation round any area is equal to the sum of the circulations round its parts, for the circulations round a common boundary are equal and of opposite sign.

Shearing Motion.—The results which we have obtained regarding pure homogeneous plane strains apply directly to this case of fluid motion. And it is easy, in addition, to deduce useful results regarding the circulation of the fluid.

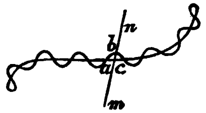


FIG. 27.

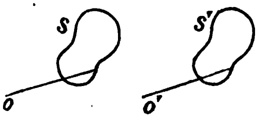


FIG. 28.

(1) *The circulation along any two plane conterminous curves, which nowhere lie at a finite distance from each other, is the same.* To see this, let abc and ac be portions of two such curves, and let

the straight line *mn* indicate the velocity *v*, which is constant along these portions provided that they are small enough. Hence the circulations round *abc* and *ac* are the products of *v* into the projections of *abc* and *ac* respectively upon *mn*. But these projections are equal, which proves the result.

(2) *The circulations round any two similarly situated and equal plane areas are equal.* For the velocities in the one curve relatively to the point *o* are equal to those in the other curve relatively to the corresponding point *o′*. Hence the circulation round *s′* differs from the circulation round *s* by the product of the relative speed of *o′* and *o* into the projection of *s′* upon the line of relative motion of *o* and *o′*. But this vanishes, since *s′* is a closed curve.

(3) *The circulation round any plane curve is proportional to its area.* For we may divide the area into a series of indefinitely small, similarly situated, and equal, parallelograms. The circulation round each of these is equal by (2). And now, since the edge of the area formed by these parallelograms is nowhere finitely apart from the given curve, the result follows from (1).

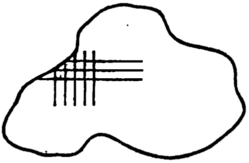


FIG. 29.

(4) *The circulation round any plane area is equal to twice the area multiplied by the angular velocity of the fluid round a perpendicular axis.* Let the area be an indefinitely small circular area. From the analogy to strain, we see that any instantaneous motion may be broken up into a pure part and a rotational part. The motion corresponding to the pure part is constant over the small area, and therefore contributes nothing to the circulation. The rotational part gives a tangential speed ωr, where ω is the angular velocity round a perpendicular axis through the centre, and r is the radius of the circle. Hence the circulation is $2\pi r \, . r\omega = 2\pi r^2\omega$. But πr^2 is the area of the circle, and so the proposition is true in this case; and, ω being constant, we see, by (3), that it holds in all cases.

Heterogeneous Motion.—A portion of the fluid may be taken so small that the motion is homogeneous throughout. The previous results are then applicable.

And, if we take any surface in the fluid bounded by a closed curve (plane or not), the circulation round that curve is equal to twice the sum, taken all over the surface, of the products of an indefinitely small portion of the surface into the angular velocity around a normal to that portion. For a sufficiently small portion of the surface may be assumed to be plane, and the result (4) applies. Hence, if a closed surface be taken, and the normals at all points be drawn outwards alone, or inwards alone, the sum of the products taken over the surface is zero. For if we draw any closed curve round the closed surface, the sum of the products, taken over each portion of the surface respectively, corresponds to equal but opposite circulation round the closed curve.

Vortex Motion.—The above conclusion may be applied to the case of a finite portion of a vortex-tube. The sides of such a tube are parallel to the axes of rotation. Therefore the ends only contribute to the integral of the angular velocity; and so the integral for each end must be equal in magnitude, and it will be of the same sign in both cases if the normals are drawn in the same direction along the tube. Hence the circulation is the same at all sections of a vortex-tube. The tube being small, the angular velocity is therefore inversely proportional to the cross-section. Hence the vortex rotates faster the thinner it is.

We conclude also that a vortex must either return into itself, forming a closed circuit, or that its ends must be at the surface of the liquid; for the velocity of rotation can neither abruptly change nor become infinite within the liquid. A smoke-ring exemplifies the former case; the eddies formed round the edge of the hand, when it is dipped into water and drawn rapidly along, illustrate the latter.

CHAPTER V.

MATTER IN MOTION.

43. *Force.*—The fundamental property of matter, which distinguishes it from the only other real thing in the universe, is *inertia.* And, in consequence of inertia, when we move a body we are conscious of making some exertion, and are accustomed to say that we exert *force.* Hence it is usual, as Newton did, to speak of force as the cause of motion; and the force may be of the nature of a push, a pull, an attraction, a repulsion, etc.

We do not yet know the nature of the physical process going on in matter which is in a state of tension, and so we figure it to ourselves by means of the mental impression caused by the muscular sense. But loudness and brightness, though they, as we shall see later, are mere subjective impressions, yet correspond to certain physical realities. We may therefore proceed to inquire whether or not there is some physical process corresponding to the impression of force.

We have already obtained (§ 7) a kinetic measure of energy; and by means of the new idea of force, we can now deduce a statical measure of it. Work is done when we move a body against the action of a force which we assert to be the cause of motion in the opposite direction. If the force is constant, we know that the work which is done is proportional to the distance through which the body is moved against the action of the force, for every equal addition to the distance is made under precisely similar circumstances. Also, we know that the amount of work which is done is proportional to the force, it being more and more difficult to produce the displacement according as the opposing force is greater. Hence, provided we define the unit of work as the work done by unit force acting through unit distance, we may write

$$w=fs,$$

where w, f, and s represent respectively the work, the force, and the distance. But the work done upon the body against the force represents (§§ 7, 8) the potential energy which the body has gained. The above equation shows that the force is the ratio of the change of energy to the space in which that change has been effected; and, if the force be not really constant, we can always choose the space so short that, throughout its description, the force is practically constant. Hence, in all cases, force is the rate per unit distance at which energy changes. The justification of the statement that we can choose the space sufficiently short is contained in the fact that we have no evidence of the existence of infinite forces in nature; so that a finite space corresponds to a finite change of energy.

Force is not conserved as energy is, although it may possess that kind of conservation spoken of in § 12. Indeed, Newton's third law of motion asserts that the total algebraic sum of the forces in the universe is zero.

And now, having arrived at a clear understanding of what 'force' really is, we may use the word, or any of its special equivalents, in subsequent sections, and speak of force as the cause of motion, without producing confusion of ideas.

44. There is no such thing in nature as a material *point*. However small a particle of matter may be, it always has a finite surface, and occupies a certain volume. And no actual force acts at a point merely, though it is often convenient to suppose that it does; it is distributed throughout a volume, or is applied over a surface. As an example of the former class of forces, we may take the force of gravitation; as an important example of the latter class, we may take the force of friction, *i.e.*, the tangential force which resists the sliding of portions of matter over each other. This tangential force is independent of the area of the surface of contact of the two bodies—so long, at least, as the surface of contact is not so small that sliding motion cannot occur without producing abrasion of the substances. It is in general proportional to the normal pressure between the bodies; but, in many substances, it depends greatly upon the time during which the contact has lasted. It may be much reduced by the use of proper lubricants.

The law of friction may be expressed by the equation

$$F=\mu' R,$$

where F is the force of friction, R is the normal pressure, and μ' is a constant, called the co-efficient of kinetic friction. When the forces which tend to produce motion are just insufficient to overcome friction, this equation becomes

$$F=\mu R,$$

where μ—called the 'co-efficient of statical friction'—is, as experiment shows, a quantity which is never less than μ'.

It follows that sliding motion will, in general, continue under the action of forces which were inadequate to start the motion.

The numbers in the second column give values of μ, those in the third give values of μ', the line of motion being parallel to the grain of the oak in the first two examples and no lubricant being employed in any case.

Wrought-iron on oak	0·62	0·62
Oak on oak	0·62	0·48
Wrought-iron on cast-iron	0·19	0·18
Cast-iron on cast-iron	0·16	0·15

45. *The Laws of Motion.*—In discussing the motion of matter, it is necessary to consider the conditions which subsist until, and must be altered before, motion is produced; that is to say, we must consider the case of rest as well as of motion. Newton's three Laws of Motion (expressed in terms of force regarded as the cause of motion) form at present the simplest foundation for the study of the phenomena of moving matter. The science which deals with these phenomena has, therefore, been called *Dynamics*, *i.e.*, the science which treats of the action of force upon matter. The unit of mass is involved in the dimensions of all dynamical quantities, in addition to the units of time and space (§ 22). It is usual to divide the subject into two parts—*Kinetics* and *Statics*—according as motion is, or is not, produced. It is impossible to doubt that ultimately a more fundamental, and at least equally simple, basis will be obtained in connection with the principles of energy. We can even at present make such a substitution for Newton's Laws, barring the simplicity. Since all motion is relative, we ought in strictness to state with reference to what origin and what directions we assert these laws; but this question is not suitable for treatment here.

46. *The First and Second Laws.*—The First Law asserts that *every body maintains its state of rest, or of constant motion in a straight line, except in so far as it is caused by force to alter that state.*

The 'rest' referred to is, of course, *relative* rest. The law includes an assertion of the conservation of the energy of the particular body considered. It implies also that a body, if moving in any path in a plane under the action of any forces, will not move out of that plane unless some other force causes it so to move. In particular, a rigid body revolving about an axis fixed in itself tends to rotate with constant angular velocity about that axis, and to keep the direction of the axis constant. For this reason we can use the (sufficiently nearly) constant motion of the earth round its axis to measure out to us equal intervals of time. In fact, the law implies the definition of equal intervals of time as times in which a body, free from the action of force, moves through equal spaces. It implies, also, that force is acting if the *velocity* changes, though the speed is constant (§ 25).

Constancy of motion in magnitude and direction does not enable us to say that no force is acting upon the body, but only that the resultant (§ 47) of all the forces is zero—that they can be combined into two equal and opposite forces. This means that the body may be simultaneously gaining and losing energy at precisely equal rates.

The Second Law gives the relation between force and the effect which it produces. It states that *change of motion is proportional to force, and is in the direction in which the force acts.* By *motion* Newton meant what is now called *momentum*—the product of the mass and the velocity of the moving body. Of course, the change of momentum is proportional to the time during which a constant force has acted, so that we may express the second law by the equation

$$ft = mv,$$

v being the change of speed produced during time t in the given mass m by the average force f. The actual force at any instant is got by making the time indefinitely small, when the equation becomes

$$f = ma,$$

a being the acceleration. And this form of the equation may be used in all cases, whatever be the value of t, provided that f and a represent the average values of the force and the acceleration during the time t. Both equations involve the assumption (or, rather, the definition) that unit force is the force which, acting upon unit mass for unit time, produces unit change of speed. This unit is termed the *absolute* unit of force, since it is independent of gravity. It is called the *poundal* if it is the force which produces an acceleration of one foot per second in a mass of one pound. It is called the *dyne* when the fundamental units are the gramme, the centimetre, and the second. From the former equation, if we suppose the force to be constant, so that the average speed produced by the force is half the final speed which it produces, we get $fs = \frac{1}{2}mv^2$, where

s is the space through which the force acts. Thus our new definition of force is equivalent to the former (§ 43).

If two forces, f_1 and f_2, act upon equal masses, the accelerations produced are proportional to those forces, for $f_1/f_2 = a_1/a_2$. Thus the Second Law of Motion gives us a method of comparing forces. It also enables us to compare masses. For, if two equal forces act upon different masses, m_1 and m_2, we have $m_1a_1 = m_2a_2$; *i.e.*, the accelerations produced are inversely proportional to the masses.

By means of the first two laws alone we can investigate the motion of a material point or of a set of disconnected particles.

For example, if g be the acceleration due to gravity, while m is the mass, the quantity mg represents the force with which the earth is attracting the mass m—*i.e.*, the weight of the body. Hence, if w be the weight, we have the equation,

$$w = mg,$$

which expresses the fundamental distinction between *weight* and *mass*. The mass m is fixed in amount; the weight w varies when g varies, and might be caused to vanish by taking the body to a region where g was zero. It is frequently convenient to use, instead of m, the equivalent quantity $V\rho$, where V is the volume of the mass m and ρ is the mass per unit volume, which is called the *density* of the substance.

Again, the equation $v = at$ (§ 26) becomes $mv = mgt$. But mv is the momentum produced in the body, which we thus see to be proportional to the time during which the body has been falling under the action of gravity. Also, instead of $v^2 = V^2 + 2as$, we have $\frac{1}{2}mv^2 = \frac{1}{2}mV^2 - mgs$, or $\frac{1}{2}m(V^2 - v^2) = mgs$, which tells us that the loss of kinetic energy is proportional to the distance through which the body has risen.

In § 29 it was shown that the acceleration of a point moving in a circle of radius r with constant speed v is v^2/r towards the centre. If the point have mass m, this corresponds to a central force mv^2/r, which, *e.g.*, in the case of a stone revolving in a sling, is supplied by the tension of the cord. The necessity for this central force gave rise to the erroneous idea of a 'centrifugal force,' which it was supposed to balance. In accordance with the First Law, the body tends to move *along a tangent*, and not from the centre, and the apparent force is really a result of inertia.

47. *Further Discussion of the Second Law.*—The above examples involve the application of a single force, constant in magnitude and direction, to a material particle. But the Second Law enables us also to investigate the motion of a material particle under the

action of any number of forces acting simultaneously, for it implicitly asserts that each force acts independently of all the others, *i.e.*, the effect produced by any force is the same as it would be if that force alone acted upon the particle when at rest.

To completely specify a force we require to know its magnitude, the direction in which it acts, and the place at which it is applied. Hence a force is a vector quantity, and so the resultant of any number of forces acting simultaneously upon a material point is to be found by the ordinary law for the composition of vectors. Indeed, this follows at once from the Second Law, since the forces are proportional to the accelerations which they produce; so, when a particle is acted upon by any number of forces, we need only consider it as moving under the action of the single resultant force.

Since a mere particle has only three degrees of freedom—all translational—three conditions completely determine its motion. These conditions might be the values of the components of the resultant force in three independent directions. In particular, when the various forces are in equilibrium, these components must vanish.

48. *Special Examples.*—We shall now apply these results to some special cases or motion of a material particle.

(1) A particle slides from rest down an inclined plane under the action of gravity. How long will it take to move over a given distance, and what will be its speed of motion when it reaches the given point?

Let m, R, and F, represent respectively the mass of the particle,

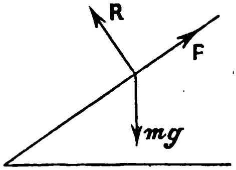


FIG. 30.

the normal pressure which the plane exerts on the particle, and the force of friction. If μ' is the coefficient of kinetic friction, we have $F=\mu'R$. We may resolve the weight of the body, mg, into its components along and perpendicular to the plane. Call these mg_1 and mg_2 respectively. Since the particle does not move perpendicularly to the plane, the force R must be equal to the component mg_2. Hence the total force acting down the plane, which is mg_1-F

or $mg_1 - \mu'R$, is equal to $m(g_1 - \mu' g_2)$. But this quantity is constant, and therefore (3) and (4) of § 26 give the required values of the time and the speed respectively, provided that in them we put $V=0$ and $a=g_1-\mu' g_2$. Otherwise, we may determine the speed by means of the equation $fs=\frac{1}{2}mv^2$ (§ 46), which directly expresses the fact of conservation of energy (work done = kinetic energy gained).

(2) A material particle is attached by an elastic cord to a point on an inclined plane down which it would slide under the action of gravity if not so attached. Find the limiting values of the tension in the cord between which motion will not occur.

The tension of the cord acts up the plane, and, if the cord be extended too far, it will pull the particle up the plane against the weight and the force of friction which always opposes the motion. If the cord be not extended far enough, the tension, even though aided now by friction, will not prevent the particle from sliding down. Let T be the tension, while the other quantities are denoted as in the preceding example. In the lowest position of the particle, the component of the weight acting down the plane must be equal to the tension minus the force of friction; in the highest position it must be equal to the tension plus the force of friction. In both cases F is equal to μmg_2, where μ is the co-efficient of statical friction since the particle is just on the point of sliding. Therefore we have $mg_1 = T \pm \mu mg_2$, from which equation we can find the two values of T.

(3) A particle of mass m is swung round in a vertical circle by means of a cord of length l. What must be its angular speed in order that the string may just be slack when the particle is at the highest point of its path?

The downward force is the weight of the particle, which must be balanced by the reaction to acceleration (the so-called 'centrifugal force,' § 46). Hence, at the highest point, $g=\omega^2 l$ where ω is the angular speed. At the lowest point the condition is $T=mg+m\omega^2 l=2mg$, where T is the tension of the cord. Hence the cord must be sufficiently strong to withstand a tension equal to twice the weight of the attached mass.

49. *Dynamical Similarity.*—We have already seen (§ 26) that, when a point moves from rest under a given acceleration, the value of the acceleration is obtained by dividing twice the distance passed over from rest by the square of the time occupied in the passage. We may therefore, neglecting the factor 2, write the expression for Newton's Second Law (§ 46) in the form

$$f=m\frac{l}{t^2},$$

where l is the distance. But in this case the sign of equality cannot be regarded as asserting numerical equality, for we have cut out a factor. It can only mean equality of *dimensions* (§§ 22, 45), and asserts that the dimensions of force involve the units of mass and length directly, and the square of the time unit inversely.

The idea of dimensions is of great importance in physics. It affords a useful check on the accuracy of algebraical work; for the dimensions of all the terms in a physical equation must be the same. But its use is not limited to this extent. For example, we may write the above equation in the form

$$\alpha f=\beta m\frac{\gamma l}{\frac{\beta\gamma}{\alpha}t_2},$$

from which we see that *if, in two similar material systems, the forces, masses, and lengths, are in the ratios* $\alpha/1$, $\beta/1$, *and* $\gamma/1$, *respectively, and if the systems begin to move in precisely similar manners, the motions will continue to be similar, provided that we compare them after the lapse of intervals of time which are in the ratio of* $\sqrt{\frac{\beta\gamma}{\alpha}}$ *to unity in the two systems.*

This principle, which was first proved by Newton, has been called the *Principle of Dynamical Similarity.* We shall subsequently get various examples of its use. (See §§ 56, 59, 100, 134.)

50. *The Third Law.*—Hitherto we have not discussed the motion of portions of matter between which there is mutual action of any kind. The first two laws of motion do not enable us to solve such problems. The requisite additional information is given by the Third Law of Motion: *The mutual actions between any two bodies are equal and oppositely directed.*

A *stress* is defined as a system of equilibrating forces, and so we may put the above law into the form: *The mutual action between any two bodies is of the nature of a stress.*

No one will question the truth of this law in the cases in which the various masses concerned are in equilibrium. Thus, when a book lies upon a table, we say that the table reacts upon the book with a pressure which is equal to its weight. But it is by no means so evident that a body, when pulled along by means of a cord, pulls backwards with a force which is precisely equal to that by which it is dragged forwards. In order to see how this can be, we must consider all the forces which are acting upon the moving body.

Let B, Fig. 31, be the body, and let F be the force with which it is pulled in the direction of F′F. Also let F″ be some other force acting upon B in the opposite direction.

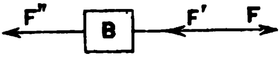


FIG. 31.

The equilibrium of B is determined solely by the equality of the forces acting upon it, *i.e.*, of the forces F and F″, and is not at all affected by the force F′ with which B reacts upon the pulling body. Hence the equality of the forces F′ and F is a matter which is entirely independent of the equality of F″ and F (that is, independent of the rest or motion of the body), and can only be proved by experiment. It is needless to add that all Newton's laws express the results of experiment or of observation.

But (as Newton himself pointed out) we may regard 'action,' not merely as force, but as the product of force into the speed which it produces in the body upon which it acts. Now, the speed produced being the (time) rate at which the force moves the body, this product is (§ 43) the (time) rate at which work is done by the force. [In modern terminology this is called the *Activity*.] Hence a second interpretation of the third law is that *the* (*time*) *rate at which a set of forces do work upon a given system is equal and opposite to the rate at which the reacting forces do work.* Had Newton been aware that heat was a form of energy, this would (see Thomson and Tait's *Elements of Natural Philosophy*) have been a complete statement of the modern principle of conservation of energy; but in his day it was supposed that work spent in overcoming friction is unavoidably and entirely lost.

Taken in conjunction with the Second Law, this law enables us to investigate the motion of bodies which impinge upon each other.

51. *Impact.*—To avoid unnecessary complications we may assume that two smooth spheres, of masses m_1 and m_2 respectively, are moving in the direction of the line joining their centres with speeds v_1 and v_2 respectively, and that after impact their velocities are v'_1 and v'_2. In most practical cases the time of impact is a very small fraction of a second, and the force is very large, so that it is impossible, without special appliances, to determine the values of these quantities. But the value of their product, called the *Impulse*, can generally be found without much difficulty. The third law tells us that the impulse is the same for each body, and hence

$$m_1(v'_1-v_1)=m_2(v_2-v'_2).$$

In addition to this we have the condition, determined experimentally by Newton,

$$v'_1 - v'_2 = e(v_2 - v_1),$$

where e is a constant, less than unity, which is called the *Coefficient of Restitution.* This condition asserts that the relative speed of separation of the two bodies is less than, but is proportional to, their relative speed of approach. It ceases to be true if the distortion produced by the impact is too great. If the bodies, after impact, move together with a common speed V, the first of these equations becomes

$$(m_1 + m_2)\mathrm{V} = m_1v_1 + m_2v_2.$$

This principle is employed in the *Ballistic Pendulum*, which is used to determine the speed of a cannon ball or of a rifle bullet. In this case the mass of the pendulum, m_2, is very large in comparison with the mass m_1 of the bullet, and v_2 is zero. The large relative value of m_2 ensures that the two masses are moving together with the common speed V before the pendulum has been sensibly deflected from the vertical. The value of V is found by observing the distance through which the pendulum swings. [From this, the height h through which the centre of inertia (§ 52) is raised is obtained, and then (§ 27) we get $\mathrm{V} = \sqrt{2gh}$.]

52. *Centre of Inertia.*—In a material system composed of masses m_1, m_2, etc., fixed relatively to each other, we can always find a point such that the product of its distance from any plane into the sum of the separate masses is equal to the sum of the products of each separate mass into its own distance from that plane. The point possessing this property is called the *Centre of Inertia* of the given set of material particles. The proof of the statement is a matter of geometry merely. Suppose that we have n points at given distances from a certain plane, and that we take another point the distance of which from that plane is the sum of the distances of the n points (say D). Now take another plane at a distance d from the former and repeat the above process with respect to it. The point so found cannot coincide with the point first found. In fact, the distance of the first found point from the second plane is $\mathrm{D}+d$, while that of the last found point is $\mathrm{D}+nd$. On the other hand, if, instead of choosing a point at a distance D from the first plane and another point at a distance $\mathrm{D}+nd$ from the second plane, we choose a point at a distance D/n from the first plane and a point at a distance $(\mathrm{D}+nd)/n$ from the second, the points so found can coincide, for their distances from the first and second

planes differ by the distance between those planes. And, since *d* may be any distance, this result holds for all planes parallel to the given one. If, for brevity, we call this point P, we see that the product of the number of the original points into the distance of P from any one of the planes is equal to the sum of the distances of the original points from that plane. Hence, if we suppose unit mass to be placed at each of the original points, P is—so far as the given set of parallel planes is concerned—the centre of inertia of these masses. Now imagine members of various independent groups of the original points at which unit masses are placed to be situated indefinitely near to each other, and we arrive ultimately at a set of distinct points with unequal masses for which—still with respect to the given set of planes—the statement made at the commencement of the section holds. The point P may of course lie anywhere on a definite plane parallel to the given set. But the process may be repeated for any other two sets of parallel planes distinct from the former set and from each other: and the point P, if it is to satisfy the given condition for each set, is then fixed. Now the given condition refers merely to space relations amongst points; and, since it holds with reference to three independent sets of non-parallel planes in space, it must, because of the essentially tri-dimensional character of space, hold with reference to all sets. Thus the statement is proved.

Since the property holds with reference to *any* plane, it holds at all instants with respect to a moving plane. Hence, instead of the words 'distance from' in the definition, we may substitute the words 'speed relatively to,' or the words 'acceleration relatively to.' The former substitution being made, the statement asserts that the momentum of the system in any given direction is equal to the momentum, in that direction, of a single mass, equal to the sum of the separate masses, moving so as always to be situated at the centre of inertia. The latter substitution being made, the statement asserts that the change of motion of the centre of inertia of any set of disconnected particles produced by the action of separate forces on the separate masses is the same as if these forces had been applied to a mass, equal to the total mass, condensed at the centre of inertia.

In consequence of the equality of action and reaction between material particles, we see that the motion of the centre of inertia of any connected set of particles is not affected by their mutual action; and that, in the case of a rigid body, we may suppose the whole mass to be condensed at the centre of inertia, and to be acted upon by the resultant force. In other words, the conditions of translational motion and equilibrium of a rigid body may be made identical with those already given in § 47 for a material particle.

53. *Moment of a Force and of Inertia.*—The *Moment of a Force* as regards rotation about an axis perpendicular to its direction is the product of the force into the shortest distance between its line of action and the axis.

A pair of parallel, equal, and oppositely directed, forces is called a *couple.* The moment of a couple about any axis perpendicular to the plane in which the forces act is equal to the product of either force into the perpendicular distance between the lines of action of the two. Let r be this distance, and let F be the common value of the forces, while P is the intersection of any perpendicular axis with

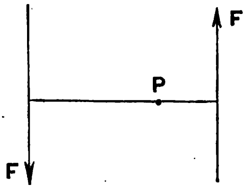


FIG. 32.

the plane in which the forces act. If x is the perpendicular distance from P to the line of action of one force, $r - x$ is the perpendicular distance from P to the line of action of the other. Hence, the sum of the moments of the forces, which is the moment of the couple, is

$$Fx + F(r - x) = Fr.$$

Now $F = ma$, where m is the mass upon which the force acts, and a is the linear acceleration, and the linear acceleration due to F is equal to the angular acceleration multiplied by r; so that if A be the angular acceleration, and we write $mr^2 = I$, we get

$$Fr = IA.$$

The quantity I is called the *Moment of Inertia* of the mass m about the given axis. The above equation corresponds in the case of rotation to the equation $F = ma$ in the case of translation. Moment of force has taken the place of force, moment of inertia has taken the place of mass, and angular acceleration has taken the place of linear acceleration.

If ω be the angular velocity, while is the whole angle through which the system has rotated under the action of the force—

supposed to be uniform—we can similarly transform the equation $Fs=\frac{1}{2}mv^2$ (§ 46). Since $s=r\theta$, and $v=r\omega$, we get

$$Fr\theta=\tfrac{1}{2}I\omega^2.$$

Hence we see that the kinetic energy of rotation acquired under the action of a given couple is proportional to the angle through which the mass has turned.

If we are not dealing with a single particle, we must write the sum $\Sigma(mr^2)$ instead of mr^2 in the above equation. But we may still put the equation in the same form as before by writing $\Sigma(m)k^2=\Sigma(mr^2)$, which is clearly allowable, since k (called the *radius of gyration*) is the only unknown quantity.

As an example, we shall investigate the motion of a cylinder rolling (not sliding) down a plane inclined to the horizon. Let r be the radius of the cylinder, and let k be its radius of gyration. If the cylinder descends through a height h, when it turns through an angle θ, and if m be its mass, the work done by gravity is mgh. This must be equal to the gain of kinetic energy. The energy in the rotational form is $\frac{1}{2}mk^2\omega^2$ where ω is the angular velocity, and that in the translational form is $\frac{1}{2}mv^2$, where v is the speed of motion down the plane. But $v=r\omega$, and therefore

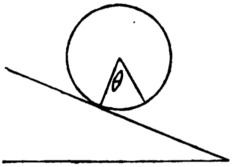


FIG. 33.

$$2gh=\frac{r^2+k^2}{r^2}v^2.$$

Had there been no rotation, we should have had $v^2=2gh$; but the speed of linear motion has been decreased because the potential energy became transformed in part into energy of rotation.

54. *Further Discussion of Moment of Inertia.*—The moment of inertia of a rigid body about any axis is equal to its moment of inertia about a parallel axis through the centre of inertia, together with the moment of inertia, about the given axis, of a mass, equal to the whole mass, condensed at the centre of inertia. For we may first suppose that the body is connected to the given axis by a rigid mass-

less rod, and that it is joined to that rod by means of a frictionless axle at right angles to the plane of rotation, and passing through the centre of inertia. In accordance with the first law of motion, the body remains at rest, so far as rotation round its centre of inertia is concerned, however rapidly it be made to revolve as a whole by means of the rigid rod. In this case, therefore, the moment of inertia of the system is equal to that of the whole mass condensed at the centre of inertia. But if now the body be rigidly fixed to the rod, the conditions remain as before with one exception—the body is now forced to rotate around its centre of inertia at the same rate as it revolves with the rigid rod. Hence the total moment of inertia has been increased by the moment of inertia about the centre of inertia.

The moment of inertia of a plane figure about an axis perpendicular to its plane is equal to the sum of its moments of inertia about any two rectangular axes lying in its own plane and passing through the point where the former axis meets the plane, for the value of r^2 at any point is equal to the sum of the squares of the distances of that point from the two rectangular axes. These two properties of the moment of inertia are of great importance in the calculation of the moment in the cases of bodies of various shapes.

55. *Rotational Equilibrium.*—There is no rotation about a given axis when Σ (Fr), taken with reference to that axis, is zero. Hence, the condition for rotational equilibrium of a rigid body is that the sum of the moments of all the forces about three non-parallel axes shall vanish. The two following examples will serve to illustrate this point.

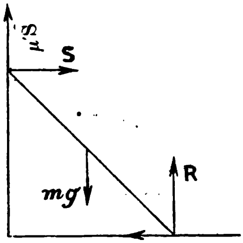


Fig. 34.

(1) A uniform ladder (Fig. 34), of length $2l$, rests, in a vertical plane, upon the ground and a vertical wall. Find the limiting position of equilibrium, the co-efficients of friction between the

ladder and the ground, and between the ladder and the wall, being μ and μ' respectively.

Under the given conditions, there is no possibility of motion except in the given vertical plane. Hence, there are only three degrees of freedom, viz., two degrees as regards translation, and one as regards rotation. The ladder will be in equilibrium, so far as translation is concerned, if the sums of the forces acting upon it in *any* two mutually perpendicular directions are zero. But it is convenient to choose those two directions which will lead to the simplest equations. We might choose the directions along and perpendicular to the length of the ladder, but each of the consequent equations would involve all the five forces which are acting. If we choose the horizontal and vertical directions, the equations involve respectively two and three forces only. Therefore, choosing the latter directions, we get,

$$S = \mu R,$$
$$mg = R + \mu' S,$$

where S and R are the normal pressures on the wall and the ground respectively.

The third relation between the quantities is obtained by equating to zero the sum of the moments of the various forces about any axis perpendicular to the plane of motion. The simplest equation is obtained by choosing the axis passing through a point (either end of the ladder), which lies on the lines of action of the greatest possible number of forces; the reason being that the moments of these forces are then zero. Let S_1, S_2, represent respectively the components of S perpendicular to, and parallel to, the length of the ladder; and let g_1 be the component of g perpendicular to it. Taking the lower end, we get

$$mg_1 = 2\,(S_1 + \mu' S_2),$$

which is independent of the length of the ladder.

If we suppose the weight and inclination of the ladder, and the values of μ and μ' to be given, we may express S_1, S_2, and g_1, in terms of S, g, and the inclination, and so we may eliminate, by means of these three equations, the quantities S and R, and obtain an equation giving the inclination in terms of known quantities.

(2) A pendulum, of length l and mass m, rotates about a vertical axis with constant angular velocity ω. Express ω in terms of g, the value of gravity, and of h, the height of the cone which the pendulum describes.

Refer to Fig. 37, p. 78. The 'centrifugal force' acts outwards horizontally, and the condition of equilibrium is that the component of this along the line of motion should be equal to the component of the weight along that line. Let r be the distance of the bob from the vertical line through the point of support of the pendulum. The centrifugal force is $m\omega^2r$. The tangential component of this is less in the ratio of h to l, and is therefore $m\omega^2rh/l$. The weight is mg, and its tangential component, being less in the ratio of r/l, is mgr/l. Equating these two quantities, we get

$$\omega^2h=g,$$

which gives the required expression.

56. *Propagation of Motion through a Non-Rigid Solid.*—As an example of the motion of a non-rigid solid we shall now discuss the problem of the passage of a wave along a stretched cord.

Let us suppose the cord to be enclosed in a smooth hollow tube, and to be drawn through it in the direction of the arrow with speed v.

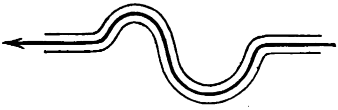


FIG. 35.

The tension T of the cord will be uniform throughout since the tube is smooth. The pressure which the cord, if not in motion, would exert upon a part of the tube where the radius of curvature has the value r is T/r per unit length. Let the cord be in contact throughout a circular arc PQ. Then, if OR bisects POQ, the resolved part of the tensions along RO is $2T.MQ/r=T.PQ/r$; for, if POQ is small, the arc PQ is equal to the chord. But the total pressure is

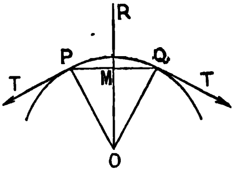


FIG. 36.

$p.PQ$, where p is the pressure per unit length of the circle. Hence $p=T/r$; that is to say, p is proportional conjointly to the tension and the curvature.

If m is the mass per unit length of the cord, mv^2/r is the 'centrifugal force' when the cord is in motion with speed v. When this is equal to T/r, *i.e.*, when $T = mv^2$, there is no pressure on the surface. The value of v which satisfies this equation is totally independent of r, the radius of curvature. Hence, when the proper speed is reached, the pressure is simultaneously taken off all parts of the smooth tube through which the cord runs; and the tube, having served the purpose for which it was used, might now be dispensed with. All parts of the cord would successively take the shape which the tube originally impressed upon the portion within it. And also, since all motion is relative, if the cord were held fixed with the given constant tension, the wave-form would run backwards along it with speed v. Hence the speed with which any disturbance will run along a cord stretched with tension T is

$$v = \sqrt{\frac{T}{m}},$$

where m is the mass per unit length.

Simple as the above proof (due to Thomson and Tait) is, the following, obtained from the principle of dynamical similarity, is at least as simple, though, on the other hand, it does not enable us to assert more than proportionality of the quantities.

The radius of curvature at similar parts of similar waves is proportional to the length l of the wave. The pressure per unit of length is therefore proportional to T/l, so that the pressure per similar length is proportional to T. Also the mass per similar length is ml, and hence we get the dimensional equation

$$T = \frac{ml^2}{t^2},$$

where t is the periodic time in which the wave length l is described. But l/t is the speed of propagation, which is therefore proportional to $\sqrt{T/m}$.

We can also investigate, by means of this principle, the law of propagation of waves through the substance of a non-rigid solid. The forces which resist similar relative motion of the parts of similar solids are proportional to the rigidities of the substances, and also to the displacements. Hence, l being length and p representing rigidity, the force is proportional to pl; and, if ρ be the density, similar masses are proportional to ρl^3. Hence the principle of dynamical similarity gives $pl = \rho l^4/t^2$; but the speed of propagation is proportional to l^2/t^2. Therefore

$$v^2 = \frac{p}{\rho}.$$

A strict investigation shows that this equation is numerically true, so that it may be used for purposes of calculation. In the case of plane distortional waves p is the distortional rigidity, formerly denoted by n; in the case of compressional waves it is the resistance to compression, formerly denoted by k; and in the case of extension of a rod, it represents Young's Modulus.

57. *Motion of a Perfect Fluid.*—A fluid may be set in motion by the action either of forces which act throughout its volume (for example, gravitational forces) or of forces which are applied to its surface (such as external pressure). A *perfect* fluid is defined as a fluid in which the pressure is always perpendicular to the surfaces of contact. It may otherwise be defined as a fluid which is entirely devoid of internal friction. Such a fluid does not exist in nature, but we may deduce various results regarding the motion of perfect fluids which will be very nearly true for actual fluids which are moving with sufficient slowness.

As an example we shall investigate the motion, under gravity, of a fluid which escapes through a small orifice in the side of a vessel, the depth of the opening below the free surface of the liquid being z.

The result may readily be deduced by considerations regarding the energy of the liquid. The kinetic energy of a quantity m of the escaping liquid is $\frac{1}{2}mv^2$. But this energy, which the escaping liquid carries away with it, is at once restored if we simply pour the liquid back again into the vessel. And the work done in raising the liquid through the height z is mgz. Hence, by the principle of conservation of energy, we have $v^2=2gz$.

In a moving fluid, the pressure is least where the speed is greatest, for at such parts potential energy has been most entirely changed into kinetic energy. Hence there is less pressure in the interior of a jet of fluid, moving through air, and therefore retarded by friction, than there is at the outside. Thus objects immersed in the fluid will be pressed inwards to the centre of the jet. This explains the support of a light body in a vertical jet of water or of air. The outward pressure upon a portion of the body in the jet is less than the atmospheric pressure on the opposite side of the body, and so the body is prevented from being dashed aside by the impact of the water.

58. *Equilibrium of a Fluid.*—When any portion of a fluid, whether perfect or not, is at rest, the resultant pressure at its boundary is always perpendicular to its surface; for, otherwise, motion would ensue.

If no external volume forces (such as gravity) act upon the fluid, the pressure per unit area of a spherical portion in the interior must have the same value at all parts of the sphere. For although, if the sphere were rigid, equilibrium would ensue if the pressures were equal at opposite extremities of any one diameter, however much they might differ for different diameters, yet, since a fluid is absolutely non-rigid, the shape of the spherical portion would alter continuously under such conditions. (Consider in illustration a hollow indiarubber ball.) And, if forces such as gravity act throughout the fluid, we may take the spherical portion so small that gravity may be neglected in comparison with the other forces; so that the statement holds in this case also. This fact is usually expressed by the phrase, 'The pressure at any point of a fluid is the same in all directions.'

From this result, so long as gravity and other volume forces do not act, we see that the pressure must be the same at all points of a fluid at rest. For we may consider *any* spherical portion, of any size, passing through a definite point, and at *all* points of all such spheres the pressure per unit area has the same value as it has at the given point; and by repetitions of this process we can reach any point of the fluid.

When gravity acts on a practically incompressible liquid the pressure per unit area increases in proportion to the depth below the surface—its increase being caused by the weight of the vertical column of the fluid standing on unit area. Hence the free surface (and every surface of constant pressure) must be a level surface. In the case of a compressible fluid, while the latter statement also holds, the pressure at any depth depends upon the law according to which the density varies with pressure. For example, in the case of air, which practically obeys Boyle's law (§ 79) the rate at which the pressure—and therefore the density—varies per unit difference of level is proportional to the density. But this is the well-known law of increase of capital at compound interest, so that the density falls off in geometrical progression as we ascend in arithmetical progression from the surface of the earth. If the density of the air were uniform throughout and equal to that at the earth's surface, the height of the atmosphere would not much exceed five miles: for the height of the water barometer is about 34 feet, and the density of water is fully 770 times that of air, so that a column of air fully 770 times higher would be required to produce the same pressure. This height is called the Height of the Homogeneous Atmosphere.

59. *Propagation of Waves in Liquids.*—We shall deal first with surface waves, and shall assume, for the sake of simplicity, that the waves are all similar, and that their ridges are parallel equi-distant

straight lines. The principle of dynamical similarity then enables us to deduce easily the law of propagation.

When the waves are propagated by gravity, the forces per similar volume are proportional to the density of the liquid, to the value of gravity, and to the square of the wave-length; and the masses are proportional to the density, and to the square of the wave-length. Hence the dimensional equation is

$$\rho l^2 g = \frac{\rho l^2 . l}{t^2},$$

which gives $$l^2/t^2 = v^2 = gl.$$

In these equations ρ represents the density of the liquid, and the other symbols have the usual significations. We see, therefore, that the speed of propagation is proportional conjointly to the square roots of the wave-length and of the acceleration due to gravity. Such waves are called *oscillatory* or *free* waves.

In the above case, it is assumed that the depth of the liquid is very large in comparison with the length of the waves. When the depth is very small in comparison with the wave-length, the above equations still apply, provided that l represents the depth of the liquid. For, when similar waves are propagated in liquids of different depths (the similarity having reference to the depth), we see that similar masses are proportional to the squares of the depths, while the ranges of vertical motion are proportional directly to the depth. Hence the speed of propagation of such waves, which are called *long* or *solitary* waves, is proportional conjointly to the square roots of the depth and of the acceleration due to gravity.

In the propagation of *ripples*, surface-tension (Chap. VIII.) is much more effective than gravitation is. If T represents the surface-tension, while l represents the wave-length, the pressure per unit area of the surface is proportional to T/l. Hence the pressure per similar area is proportional to T, for we are not concerned with lengths measured parallel to the ridges of the waves. The similar masses are proportional to ρ and to l^2, and so we get

$$v^2 = \frac{\mathrm{T}}{\rho l}.$$

The speed of propagation of a ripple is therefore proportional directly to the square root of the surface-tension, and inversely to the square root of the product of the density into the wave-length. Thus we see that ripples run faster the smaller they are, while oscillatory waves run faster the larger they are. Hence there is a certain size of wave (about two-thirds of an inch in length in the case of water) which runs slowest. Smaller waves run more quickly because the effect of surface-tension preponderates; larger waves run more quickly because of the increased effect of gravity.

The distortional rigidity of most fluids is zero, and so distortional waves cannot be propagated; but the result at the end of § 56 applies to the case of compressional waves in fluids.

CHAPTER VI.

PROPERTIES OF MATTER.

60. *Definitions of Matter.*—It is only possible at present to give provisional definitions of matter; for we cannot yet say, possibly may never be able to say, what matter really is. It may be defined in terms of any of its distinctive characteristics. We may say that *Matter is that which possesses Inertia.* Or again, since we have no knowledge of energy except in association with matter, we may assert that *Matter is the Vehicle of Energy.* Another statement (which, from the results of Chap. I. and § 43, we see to be a different form of the latter) is that 'matter is that which exerts, or can be acted upon, by force.' Further knowledge would probably make it evident that all three definitions are merely differently worded statements of the same fact.

61. *States of Matter.*—Matter is usually spoken of as existing in three different states—the solid, the liquid, and the gaseous.

A portion of matter in the solid state possesses a definite form of its own, and considerable force has to be applied in order to produce an appreciable change in the form. When in the liquid state, matter, on the other hand, possesses no definite form of its own, and can be made by application of the slightest force to change whatever form it happens to have. A similar statement holds in the case of a gas or vapour. But a gaseous body differs from a liquid in that its volume is limited only by the volume of the closed vessel in which it is contained; while the volume of a given quantity of liquid is perfectly definite, under given physical conditions, however large the containing vessel may be.

Still—although these rules are of general applicability—it must not be supposed that there is any hard and fast distinction between a solid and a liquid, or between a liquid and a vapour: possibly (but there is no experimental proof of the truth of this statement) there may be, under certain physical conditions, no sharp line of demarcation between the solid and the vaporous states of matter.

As the temperature of sealing-wax is gradually raised, the substance slowly passes from the solid into the liquid condition, and, for some time, we cannot strictly call it either a liquid or a solid. The transition from ice to water seems to occur suddenly, but analogy would indicate that the process is really a continuous one. We shall also find (Chap. XXI.) that the passage from the vaporous to the liquid condition may be made without break of continuity. A very notable and extremely important example of the non-rigidity of the above distinctions occurs in the case of shoemakers' wax. This substance so far resembles a brittle solid that it will break into splinters under the blow of a hammer; and yet, under the action of slight long-continued forces, it can be moulded into any shape we please.

62. *General Properties.*—Certain properties are common to all portions of matter in whatever state or physical condition they may be, and are, therefore, called *general* properties.

Chief among them is the already-mentioned property of *inertia* (§ 6). Because of it air-currents are effective in propelling ships, earthquakes throw down massive erections, the sea continues to be agitated after a storm ceases, a cannon ball has destructive power, and so on. Many mechanical applications of the property are made. Thus a fly-wheel is of advantage in steadying the motion of an engine in virtue of its being inert; and a hole can be punched through a thick plate of iron by a single stroke when the drill has sufficient inertia. We take the inertia of a body as the measure of its *mass*, *i.e.*, of the quantity of matter which it contains.

Again, all matter occupies space; and we consequently say that it has the property of *extension*. But the occupancy of space further involves the idea of form, and so we recognise *form* as a property of matter. There is little more to be said on this point except in the case of the form of crystallised bodies, to the consideration of which a considerable part of Chap. VIII. will be devoted.

So far as we know, any one portion of matter occupies a given portion of space to the utter exclusion of all other matter. This certainly holds in the case of any visible portion. Hence we look upon *impenetrability* as a general property of matter. But the quality of impenetrability does not interfere with interpenetration of matter, the possibility of which depends upon the existence of pores in any finite portion of matter. The corresponding property is called *porosity*. All matter is more or less sponge-like in structure, the space in the so-called 'internal' pores being in reality external to the material of the body. In the case of substances such as cork, wood, coarse sandstone, etc., the porosity is very evident. The porosity of metals is shown by the fact that gases can pass through them. Thus palladium has a remarkable power of absorbing or 'occluding' hydrogen; carbonic oxide passes readily through red-hot iron; and gases formed by the decomposition of electrolytes pass through the metallic electrodes. Bichromate of potassium passes into the pores of glazed earthenware, and, crystallizing inside, gradually breaks up the substance. The porosity of liquids is evidenced by their absorption of gases.

Vitreous bodies alone have not yet been directly shown to be porous; but we may fairly conclude that we have not yet found the proper method of testing the point, and that, the proper method being found, they too will prove to be no exception to the rule.

A very noteworthy example of interpenetration occurs in the alloying of certain metals, such as tin and copper. The bulk of this alloy is considerably less than the sum of the bulks of its constituents. This phenomenon does not occur when gold and silver are alloyed, and for this reason only was Archimedes' famous test of the impurity of Hiero's crown conclusive. A certain weight of pure gold had been given to a smith for the purpose of making the crown; but it was suspected that he had abstracted some of the gold, replacing it by an equal weight of silver. Archimedes knew that, weight for weight, silver is bulkier than gold; and hence he concluded that the crown would be bulkier than the given amount of pure gold if silver had been used as an alloy. The problem which he required to solve was therefore the determination of the bulk of a solid which was so irregular in shape that no method of estimation by direct measurement was applicable. He determined this by measuring the volume of the water which the crown displaced. Had contraction taken place, the bulk and weight of the alloy might have been the same as those of the pure gold.

Another property of all matter is *divisibility*. The question of the infinite divisibility of matter will be further alluded to in the chapter on the constitution of matter. In the meantime we are only concerned with examples of extreme division. Many such occur readily. Films of gold and of other metals may be made so thin as to be transparent. A film of gold precipitated by chemical means and burnished so that it forms a continuous sheet may be of no greater thickness than one ten-millionth part of an inch. A quartz-fibre may be made so fine as to be utterly invisible. The vapour from a particle of sodium will tinge a flame continuously of a deep orange colour for hours at a time. A single drop of a strongly coloured liquid will continuously tinge a very large quantity of water, and its presence may be made evident by chemical means long after the eye ceases to detect it. Further examples of the extreme smallness of portions of matter will be given in the chapter on the constitution of matter.

All matter is capable of having its volume diminished under pressure to a greater or less extent, and so we speak of its *compressibility*. This subject will receive detailed treatment in Chaps. VII., VIII., and IX.

Again, all matter is deformable. But it is often more convenient to speak of its *rigidity* than of its deformability, that is, the property in virtue of which it resists deformation. This property, too, will be discussed subsequently.

We shall also afterwards consider more specially *elasticity*, which, in one or other of its two forms, exists in all kinds of matter; and *viscosity*, that is, the property in virtue of which there is resistance to relative motion of the particles of a body.

Gravitational attraction is another general property of matter. According to Newton's Law, 'every particle of matter in the universe attracts every other.'

63. *Special Properties.*—Many other properties might be enumerated which are conspicuously present in some substances, and are as conspicuously absent from others, such as *plasticity*, *ductility*, *brittleness*, *tenacity*, etc. Again, many properties refer to matter in connection with special forms of energy; for example, *dispersive power*, *thermal* and *electric conductivity*, *magnetic permeability*, *translucency*, *opacity*, etc. Indeed, all the properties of matter might be naturally investigated in a treatise on energy; for we have no notion of what might be the properties of matter devoid of energy. Most probably it would not be matter at all in the sense in which we use the word.

64. *Contingent Properties.*—The manifestation of certain properties often depends upon conditions which may be termed accidental. Hence these are frequently called contingent properties. For example, the manifestation of brittleness by water-substance depends upon its thermal condition. Again, the *weight* of a body is contingent, depending for its existence upon the presence of a second body, and varying in amount with the distance between the two. A body placed in a cavity at the earth's centre would possess no weight (§ 72).

65. *Specific Properties.*—Many of the properties of a body depend upon the size of that body. Thus a portion of a given substance is

more massive than another portion of the same substance in proportion as its bulk is greater than the bulk of that other.

It is frequently very essential to define a property of a body in such a way as to make it independent of the size of any particular specimen. For example, the *density* (specific mass) of a substance is the mass per unit volume of that substance; the *specific gravity* is the weight of unit volume, expressed in terms of that of water as the standard; the *specific weight* might be defined as the weight per unit volume, expressed in absolute units. We also define *rigidity*, *viscosity*, etc. (§§ 84, 104), as specific properties.

66. *Molecules and Atoms.*—A visible portion of any chemically compound substance, *e.g.*, water, may be divided into smaller parts, each of which has the same chemical peculiarities as the whole had. These parts may even be so small as to be invisible to the most powerful microscope. But at last a stage is reached where further division cannot occur without the production of substances which are chemically different from the original one. These smallest parts, which are chemically similar to the whole, are called *molecules.*

Every molecule can be divided further into what are termed its constituent *atoms.* These atoms are dissimilar when the molecule is complex like that of water: they are all precisely similar in a simple molecule, such as that of hydrogen. An atom is an indivisible part—that is, indivisible by any means at present at our disposal. In all probability, it is really compound.

A substance whose molecules are composed entirely of similar atoms is called an 'elementary' substance, and the kind of matter composing it is called a 'chemical element.' We know of the existence in nature of only a comparatively small number of elements. This suggests that all the so-called elementary atoms may be in reality compound (or that at least all but one may be so) being formed from a single truly elementary type. This hypothesis was first advanced by Prout, and is in accordance with the unity which is everywhere observable throughout nature.

67. *Molecular Forces.*—A considerable, frequently a very great, amount of work is necessary in order to separate the molecules of a body. When separated, we are accustomed to say that they possess potential energy of molecular separation—the increase of potential energy being equivalent to the work spent in producing the separation. It is also customary to say that *force* is required to effect the separation, and to speak of the molecules as being held together by *molecular forces* which tend to maintain them in their relative positions against the action of external force.

The work done in separating molecules beyond the range of their mutual forces is usually great, and practically the whole of it is performed in an excessively short distance. Hence it is said that the molecular forces are extremely powerful, but that they are insensible at sensible distances. In confirmation of this we may firmly press together two leaden bullets at a part where their surfaces have been freshly cleaned. They will cling together so that the lower one may be lifted by means of the upper, the magnitude of the so-called molecular forces near the surface of contact being sufficient to overcome the gravitational attraction of the whole earth. If there is a film of oxide on the metal the experiment will not succeed, the film having too great thickness to allow of appreciable molecular attraction across it. (For further statements on this subject, see Chap. VIII.)

Many of the properties of matter may be said to depend on the nature of the molecular forces. *Tenacity* depends upon the extent to which these forces can overcome external forces which tend to draw the molecules apart. *Malleability* is a property essentially analogous to the preceding. It is the property in virtue of which a substance may be extended in two directions, while it is contracted in a direction perpendicular to these by the application of great pressure, its volume remaining practically unaltered. In testing tenacity the extension occurs in one direction only, and there is contraction in all directions perpendicular to that one. The *brittleness* of a body is due to the comparative ease with which the molecules can be separated beyond the range at which their mutual forces are appreciable. *Rigidity* depends upon the ability of the molecular forces to resist alteration of the relative positions of the molecules of a body. We therefore recognise two kinds of rigidity —rigidity as regards bulk, and rigidity as regards form. *Viscosity* depends upon the ability of the forces to resist shearing motion.

68. In the immediately succeeding chapters, a more detailed examination of some of the most important of the properties alluded to above will be given. Others will be treated as occasion may arise subsequently. Gravitation will be discussed further in the remaining sections of the present chapter.

69. *Gravitation.*—All bodies in the neighbourhood of the earth's surface possess potential energy, which, when circumstances permit, is invariably changed into kinetic energy of motion *towards* the earth. Hence we say that each body is 'attracted' to the earth with a force which is termed its 'weight.'

Bodies made of the same material may be roughly judged to be heavy in proportion to their bulk, *i.e.*, in proportion to the quantity of matter contained in them. But we need not rest content with a roughly approximate rule, for the law that weight is proportional to mass is capable of as rigid proof as can be given by the most accurate physical methods.

In the first place, every body—except in so far as the resistance of the air is concerned—takes the same time to fall through a given distance. In other words, the acceleration of motion is the same in all cases. Hence, by Newton's Second Law, the force (weight) is proportional to the mass.

The fact that the time of oscillation of a simple pendulum is independent of the mass of the bob furnishes a more readily obtainable proof of the proportionality of weight and mass. For, when the oscillations are small, the bob moves through a horizontal range which is large in comparison with the vertical range through which it simultaneously falls. There being no horizontal acceleration, the time of fall through the given vertical distance is made large. Now the weight of the bob may be resolved into its components along the string and perpendicular to it. The latter component, being in the direction of motion, is the force which makes the pendulum oscillate. We may represent the weight by the vertical line (Fig. 37) corresponding to the undisturbed position of the string, in which case the force producing oscillation is represented by the perpendicular, from the undisturbed position of the bob, upon the inclined line. But, when the oscillations are very small, this perpendicular is practically equal in length to the arc through which the bob is displaced. Hence (§ 33) the time of all small oscillations is constant. Therefore, in addition to the already mentioned advantage that this time is comparatively large, there is the additional advantage that we can, as it were, magnify it further by frequent repetitions, and so may obtain a very accurate measure

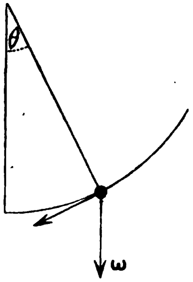


FIG. 37.

of it—a much more accurate measurement than any which could be obtained directly.

Again, a body has the same weight whether its surface is large or small, and whether it is in a single lump or broken up into parts. This proves that the outer parts do not screen the interior parts from the action of gravitation.

70. *Specific Gravity.*—We have already defined density or specific mass as the mass contained in unit volume, and specific gravity as the weight of the mass per unit volume measured in terms of the weight of water per unit volume. To make the latter statement strict we must state the pressure and temperature of the water. It is usual to take water at 4° C., and under the pressure of one atmosphere. Of course, the actual pressure of the atmosphere varies, but *one atmosphere* is defined as a pressure equal to the weight, at 0° C., of a column of mercury 760 millimetres high and of unit section. Since weight is proportional to mass, tables of specific gravity will also be tables of density, provided only that we take, as the unit of mass, the mass of unit volume of that substance which is taken as the standard for specific gravities. On the French system the unit of mass (the gramme) is the mass of water, at 0° C., and a pressure of one atmosphere, which is contained in one cubic centimetre.

If a solid cylinder be immersed vertically in a liquid, the downward pressure on the upper end is less than the upward pressure on the lower end by the weight of the liquid displaced (§ 58). And this statement holds even if the ends of the cylinder are not horizontal planes, for the horizontal component of the pressure is of no account in the present question. Hence any solid, which is denser than water, weighs less when weighed in water than in vacuo by the weight of water which it displaces, so that, if we take water under the standard conditions (or under any known conditions, as it is easy to make the necessary reductions) we can find the weight of an amount of it which is equal in bulk to a given solid by subtracting the weight of the solid in water from its weight in vacuo (or, practically, in air). From this we get the specific gravity of the solid by dividing its weight by the weight of the equal bulk of water. Various modifications of this method are used for special objects. A solid which is lighter than water may be fastened to a heavier body which makes it sink. In this way the weight of water displaced by the two together may be found, and the weight displaced by the heavier body is found by the previous process. The difference of these is the weight of water displaced by the lighter solid.

In the case of a liquid we may obtain the specific gravity by weighing a solid successively in air, in the liquid, and in water, The difference of the first and second weights divided by the difference of the first and third is the specific gravity of the liquid.

In another method hollow glass balls (*Specific Gravity Beads*) are used. The specific gravity of the liquid is equal to that of one of the balls which neither floats nor sinks in the liquid. The fact that the weight of liquid displaced by a floating body is equal to the weight of that body constitutes the principle upon which the ordinary *Hydrometer* is constructed. This instrument consists of a hollow glass ball (weighted with mercury) and a graduated stem. The graduations may indicate the volume of the immersed portion. The quotient of the weight of the hydrometer by this volume gives the required specific gravity. More conveniently, the specific gravity may be directly indicated on the stem.

The determination of the specific gravity of a gas is a matter of great experimental difficulty, though relative values for various gases may be obtained with comparative ease. The quantity sought is the weight of the gas which fills a vessel of known volume. When the gas is exhausted for the purpose of weighing the vessel alone, the volume of the vessel alters, because the gaseous pressure is removed from the interior, and careful corrections must be applied.

The following table gives values of the densities or specific gravities of various substances:

Platinum	21·2	Mercury	13·59	Chlorine	2·450
Gold	19·3	Bromine	2·98	Carbonic acid	1·524
Lead	11·3	Sulphuric acid	1·84	Hydrochloric acid	1·270
Silver	10·5	Carbon bisulphide	1·27	Oxygen	1·106
Copper	8·9	Acetic acid	1·06	Nitrogen	0·972
Iron	7·7	Turpentine	0·86	Carbonic oxide	0·968
Zinc	6·9	Alcohol	0·80	Ammonia	0·590
Aluminium	2·6	Naphtha	0·80	Marsh gas	0·558
Lithium	0·58	Ether	0·72	Hydrogen	0·069

The densities of the solids and liquids are given in terms of the density of water under the standard conditions. The densities of the gases are given in terms of the density of air as unity. In terms of water as the standard the density of air is 0·00129 approximately. The density of a solid varies somewhat with the mechanical treatment to which it is subjected.

71. Assuming the truth of Kepler's Laws regarding the motion of a planet, Newton deduced certain results with respect to the nature

of attraction between the planet and the sun which would account for these laws. The laws are:

I. Each planet moves in an ellipse of which the sun occupies one focus. (The path of a comet may be any conic section.)

II. In that ellipse the radius-vector traces out equal areas in equal times.

III. The square of the periodic time is proportional to the cube of the mean distance between the planet and the sun.

From the second of these Newton showed that the attraction must be in the line joining the two bodies. He then showed from the first two laws that it must vary inversely as the square of the distance between the bodies. The third law, from which the latter result might have been independently deduced, was then used to indicate that gravity depends only on the quantity, and not on the quality, of matter; for it shows that the ratio of mass to attraction would be constant in the cases of all the planets if they were placed at a common distance from the sun.

If, from the above evidence, we now assert Newton's great Law of Gravitation that *Every particle of matter in the universe attracts every other particle with a force whose direction is that of the line joining the two, and whose magnitude is directly as the product of their masses, and inversely as the square of their distance from each other*, we see that Kepler's Laws cannot be strictly true. In the first place, all the bodies in the solar system (including the sun) will revolve about the centre of inertia of the system, which is not necessarily nor actually situated at the sun's centre; and again, their paths cannot be true ellipses because of mutual attraction.

But we *know* that Kepler's Laws are not strictly true; and, further, the deviations from them are precisely such as should result from mutual attraction amongst the various planets. Indeed, by assuming the truth of Newton's Laws of Motion and of the Law of Gravitation, Adams and Leverrier were able to predict independently the existence and position of the previously unknown planet Uranus.

The law of gravitation is supported by almost as strong proof as any theoretical statement can possibly have.

72. So far as we have gone we have looked upon the sun and the planets as mere material points. The justification of this depends in part on the smallness of their linear dimensions in comparison with their distance asunder, but is more fully contained in the latter of two theorems due to Newton: 1°, *A thin spherical shell composed of uniform gravitating matter exerts no resultant attraction*

upon a particle in its interior; and, 2°, *it attracts an external particle as if its whole mass were condensed at its centre.*

To prove the former statement, consider a cone of very small angle with its vertex at the given internal point. The little, practically plane, areas which this cone cuts out from the shell are equally inclined to its axis, and so are proportional to the squares of their distances from the vertex; but their masses are in the same proportion. Therefore their attractions upon a particle at the vertex are equal and opposite; and this is true for all such pairs of areas of which the whole shell is made up. An equally simple proof of the second statement will be found in the course of the discussion of tubes of force in Chap. XXIV.

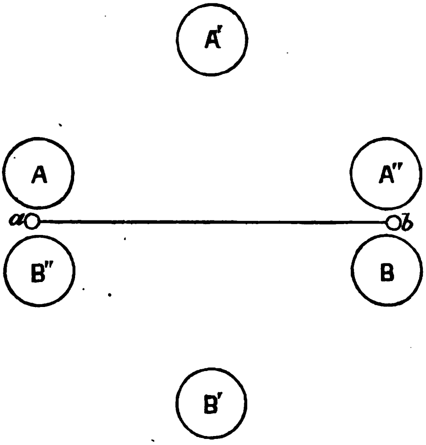


FIG. 38.

Since the proposition is true of a shell, it is also true of a solid sphere which may be built up of a number of thin concentric shells.

The above case furnishes one example of that limited class of bodies which attract, and are attracted by, external bodies, as if their whole mass were condensed at a definite point, called their *centre of gravity.* Other examples will be given when we discuss electrical images. The centre of gravity, when it exists, always coincides with the centre of inertia.

73. The second theorem is made use of in Cavendish's method of determining the mass (and, consequently, the mean density) of the earth.

Two small leaden balls, of mass m, are attached to a light rigid rod or tube, ab, which is attached at its middle point to a vertical wire. The couple required to twist the wire through a given angle is determined by observations upon the time of oscillation of the system. Two large leaden balls of mass M, which were originally in the positions A′ B′, are placed in the positions A, B. The mutual attractions of the balls A and a, B and b, deflect ab from its normal position, and it oscillates about a new position. The angle through which ab is deflected is determined by means of a beam of light reflected from a mirror which is fastened to the suspending wire. Similar observations are made with the large balls once more in the positions A′, B′, and finally in the positions A″, B″, at the same distance as before from a and b, but on their other sides, so as to deflect ab in the opposite sense. The mean of the deflections is taken, and the couple required to produce it is known. In this way the absolute value of the attraction between the two masses M and m, their centres being at a given distance apart, is found; and it may be compared with the attraction of the earth upon the small ball.

Professor C. V. Boys has recently succeeded in drawing out extremely fine fibres of quartz, the torsional rigidity of which is so small that, by their means, the mutual attraction between two lead pellets can easily be made manifest.

74. The *Schehallien* experiment, undertaken with the object of determining the value of the mean density of the earth, was of precisely the same nature. The deflection of the bob of a pendulum from the true vertical under the attraction of the mountain Schehallien was obtained by determining the apparent difference of latitude of two places—one being situated to the north, and the other to the south of the mountain—by means of a pendulum used as a plumb-line. If from this we subtract the true difference of latitude of the places, we get a measure of the attraction of the mountain upon the bob of the pendulum as compared with the attraction of the earth upon it; for the attraction of the mountain pulls the pendulum from the vertical in such a way as to increase the apparent latitude of the northern point, and to diminish that of the southern. The great objection to this method lies in the fact that our knowledge of the mass of the mountain is necessarily very imperfect.

Yet another method consists in observing the times of oscillation of pendulums of the same length, one of which is situated at the top, and the other at the bottom, of the shaft of a coal-pit. In this way we get a comparison of the attractions of the whole earth, and of the earth minus a shell of thickness equal to the depth of the mine, upon a mass situated at known distances from the centre. Here, again, uncertainty arises, because of our limited knowledge of the density of the earth's crust, which must be assumed to be equal to that in the neighbourhood of the shaft. This is called the *Harton* experiment, because it was performed at the Harton mines.

The value of the mean density of the earth obtained by the Schehallien experiment is considerably less than, and the value obtained by the Harton method is considerably greater than, the mean of the (very closely accordant) results obtained by different experimenters who used the Cavendish method. Their results show that the mean density is about 5·5 times the density of water.

Poynting has recently carried out experiments in which the attraction of the earth upon a known mass is measured by a beam-balance method.

Hypotheses Framed to Explain Gravitation.

75. Various attempts at an explanation of gravitation have been made—all more or less unsatisfactory.

One of the most noted of these is Le Sage's hypothesis of ultramundane corpuscles. According to Le Sage, gravitation is due to the bombardment of bodies by numberless small material particles which are darting about in space with great speed in all directions. A single material body placed in space would not be impelled in any one direction more than in another, for the corpuscles would batter it equally on all sides. But two bodies in space would be driven together—provided, at least, that their distance apart were small in comparison with the free-path (§ 124) of the corpuscles—for those sides of the bodies which face each other would be shielded to a greater or less extent from the bombardment. If the dimensions of the bodies were very much smaller than their distance apart, the force of attraction (the incongruity of which term is evident from the fact that the force is here really one of *impulsion*) would vary in direct proportion to the cross-sectional area of each of the two bodies as seen from the other. But this is not the gravitational law. In order to obtain it, we must make the supposition that the molecules of matter are so far apart that the number of corpuscles which pass completely through (say) the earth, without striking any part of it, is enormously larger than the number of those which are stopped by it. In this case, every molecule in the interior will be bombarded equally with an exterior particle. The force also will be inversely as the square of the distance, and so we get Newton's Law. In order that the planets should not experience appreciable resistance to their motion around the sun, it is necessary to assume, further, that the speed of the planets is exceedingly small in relation to that of the corpuscles.

76. The law of gravitation may be worked out into all its consequences (at least, so far as our methods avail) without any knowledge of the mechanism by which it occurs. We require merely to assume that it acts directly at a distance. But, as Newton remarked, no one competent to think correctly on physical matters will be content with this assumption. He himself suggested the rarefaction of the ether in the neighbourhood of dense bodies as a possible explanation. Lord Kelvin has pointed out a dynamical method of producing this diminution of pressure. An incompressible fluid, filling all space, which is brought into existence, or is annihilated, at the surface of every particle of matter at a rate proportioned to the mass of that particle, and which is annihilated, or produced, at an infinite distance, at the same total rate would supply the necessary means. (See § 57.)

Waves traversing a medium would have the effect of making bodies immersed in it approach each other.

The property of *dilatancy* (Chap. XXXII.) in a medium composed of rigid particles in mutual contact would also account for a gravitational action on bodies placed in it.

A certain stress in Maxwell's electro-magnetic medium (Chap. XXX.) would account for it too. So, as we have seen, would Le Sage's ultra-mundane corpuscles. But, to every provisional hypothesis yet brought forward, some objection, more or less conclusive, may be advanced.

If gravitation be due to action propagated through a material medium, its propagation through finite distances must occupy a finite time. We can assert merely that the speed of propagation is large in comparison with planetary velocities; for no such modification of the planetary motions as would be entailed were it otherwise is observable.

The Nebular Hypothesis.

77. Various nebulæ, when examined under sufficiently high magnifying power, are seen to be merely groups of stars like our own stellar system. But others cannot be so resolved; and, among them, there is great variety of constitution. Some appear to be comparatively uniform throughout, while others seem to vary greatly in density at different parts. Others, again, have very dense nuclei with a faint nebulous surrounding.

Such facts as these suggested to Laplace his famous Nebular

Hypothesis. According to this hypothesis, we see in these nebulæ solar systems, such as our own, in various stages of formation. The sun was formed by the mutual gravitation of its parts from a state of diffusion throughout space. And, to account for its rotation, these parts must have had the same amount of moment of momentum about an axis as that which the sun at present has. In all probability they had small translational velocity before they began to gravitate towards each other—small, that is, relatively to the speed which they would acquire under the action of gravitation ; for, if this were not so the chance of their colliding would be vanishingly small. The collision of these parts would produce great heat, which would result in the production of a nebula, extending beyond the orbit of Neptune, and slowly revolving on its axis. As the nebula gradually shrank, its angular velocity would increase until, by 'centrifugal force,' a ring of matter was left behind as the body of the nebula still farther shrank. The breaking up of this ring (which would usually occur, as dynamical principles show), and the subsequent agglomeration of its parts, would result in the formation of a planet. And so the development would proceed.

Laplace's hypothesis cannot, as he originally stated it, give a *full* account of all the phenomena of our solar system ; but this we do know, that some such hypothesis must be true if our sun's heat has had a physical origin such as our present knowledge is adequate to explain. We cannot conceive of any store of energy, sufficient to account for the immense radiation of heat from the sun, except the potential energy of separated masses of gravitating matter.

CHAPTER VII.

PROPERTIES OF GASES.

78. *Compressibility.*—Throughout this investigation we shall assume that the temperature of the gas remains constant. The effects which result from changes of temperature will be more conveniently treated in the chapter on the effects of heat.

All gases are compressible; that is, their volume can be diminished by the application of pressure. We shall see afterwards that sound could not pass at a finite rate through a gas which was not compressible. So that the mere fact that gases can convey sound constitutes a proof of their compressibility.

79. *Boyle's Law.*—The law which very completely, though not with absolute accuracy, represents the relation between the pressure and the volume of air (and many other gases) was discovered experimentally by Boyle, who showed that *the density of a gas is directly proportional to the pressure.* In symbols, ρ being the density and p the pressure, this is

$$\rho = cp,$$

c being a constant. The density—that is, the mass or quantity of matter in unit volume—is numerically equal to the reciprocal of the volume containing unit mass. Hence, v being this volume, we may write, instead of the above,

$$pv = c,$$

where the constant c is the reciprocal of the former one; or, in words, *the volume is inversely proportional to the pressure.*

Boyle's apparatus consisted of a glass U-tube (Fig. 39) with a long and a short limb. The long limb was open to the atmosphere, while the short one was closed, and contained a quantity of air, which was separated, by means of mercury filling the bend of the tube, from the outside air. The level of the mercury was the same in both limbs of the tube, and so the enclosed air (the volume of which was carefully noted) was at atmospheric pressure (§ 58). Mercury was then poured into the open limb, until a difference of level equal to the height of the mercury barometer was established. The air inside was therefore under a pressure of two atmospheres, and its volume was found to have been halved; and so on with other values of the pressure.

A slight modification of the apparatus enables us to prove the law under diminished pressure. AB (Fig. 40) is a vessel containing mercury. The glass tube *ab*, which is closed at the end *a*, but is open at *b*, is filled with mercury, and inverted in AB. Being shorter than the height of the mercury barometer, the tube *ab* remains filled. Air or any other gas may now be introduced into it until the mercury inside is at the same level as that outside. Under these conditions the gas in *ab* is under atmospheric pressure. If *ab* be raised, the mercury in it stands at a higher level than that outside, and the gas expands, since it is under diminished pressure.

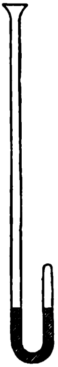


FIG. 39.

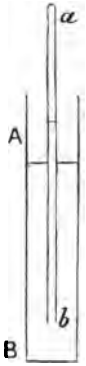


FIG. 40.

The ideal gas which rigidly obeys Boyle's Law is called a *perfect gas*.

80. *Compressibility of a Perfect Gas.*—One gas is more compressible than another in direct proportion to the alteration of volume produced by a given pressure, and in inverse proportion to the pressure required to change the volume to a given extent. Hence we measure the compressibility by the ratio of the fractional change of volume—called the *compression*—to the change of pressure which produces it. That is to say, if the volume V changes by the quantity v when the pressure alters by the amount p, the compressibility is measured by the ratio v/Vp.

By Boyle's Law we have

$$PV = c,$$

and

$$(P + p)(V - v) = c.$$

Therefore

$$pV - Pv = 0,$$

since we can neglect pv, which is the product of two small quantities. This gives

$$\frac{v}{Vp} = \frac{1}{P},$$

that is, *the compressibility of a perfect gas is inversely proportional to the pressure.*

81. *Deviations from Boyle's Law.*—Though no gas is perfect.

yet many gases do not greatly deviate from Boyle's Law throughout a considerable range of pressure.

Air is more compressed than it should be in accordance with the law until a pressure of nearly one ton's weight on the square inch is reached. After this point, the compression is *less* than the calculated value. A reason for this is simply that the volume of the gas is not capable of indefinite decrease, while Boyle's Law asserts that under infinite pressure the volume will become zero.

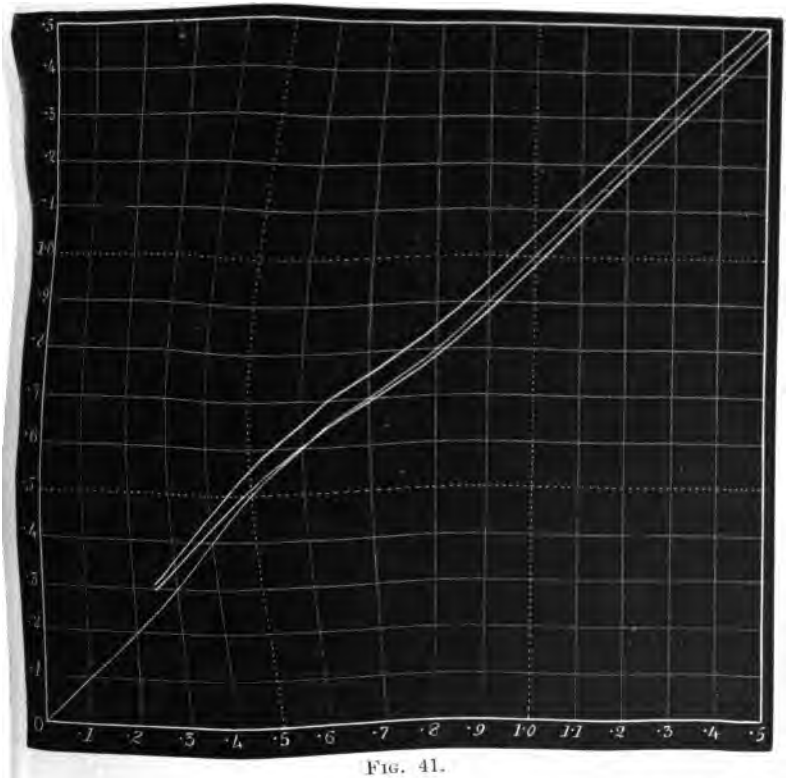


Fig. 41.

Hydrogen, unlike air, is, at ordinary temperatures, always less compressible than the law indicates. Nitrogen, along with many other gases, resembles air.

These results are exhibited graphically in Fig. 41. The actual volume of the gas is measured along the vertical axis, while the

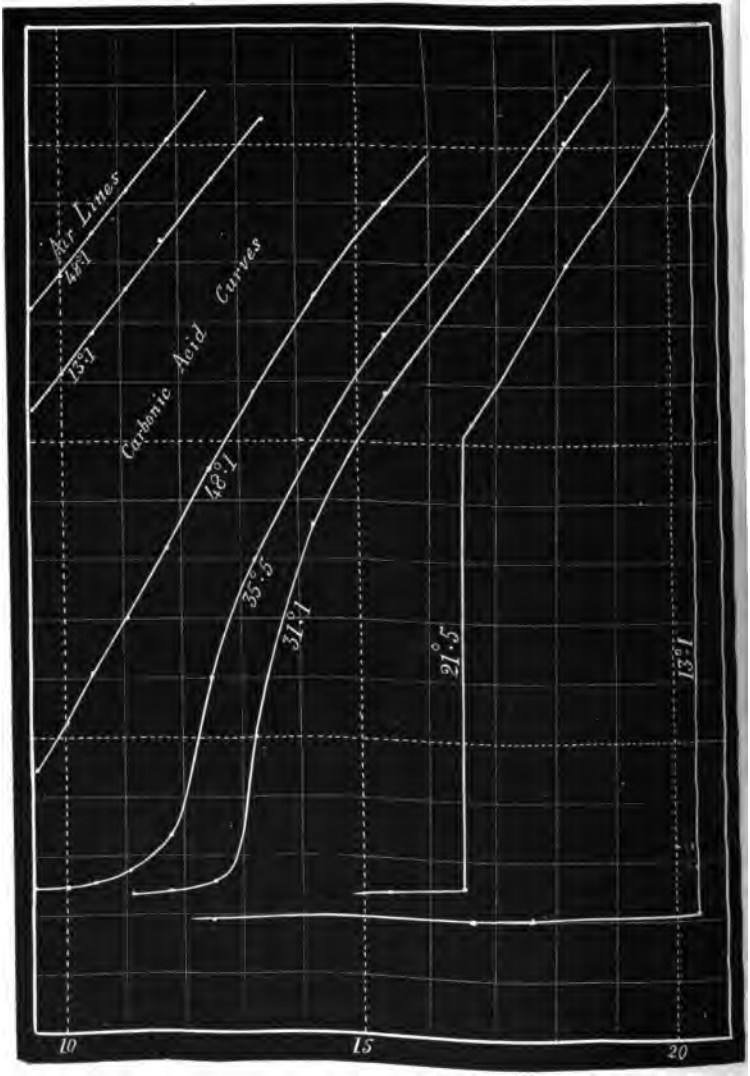

Air Lines
48°·1
13°·1
Carbonic Acid Curves
48°·1
35°·5
31°·1
21°·5
13°·1
10
15
20

FIG. 42.

volume of a perfect gas under the same pressure is measured along the horizontal axis. The straight line passing upwards through the origin at an angle of 45° is obviously the graph for a perfect gas. The curved line intersecting the perfect gas-line, and, like it, sloping upwards towards the right, represents the action of air; and another curve, sloping similarly, is the graph for hydrogen.

82. *Compression of Vapours.*—A vapour, though it may obey Boyle's Law throughout a considerable range of pressure, ultimately deviates more and more from that law as the pressure rises. The direction of the deviation is similar to that of air at pressures less than 152 atmospheres, *i.e.*, the vapour is more compressed than a perfect gas would be. When the pressure has become sufficiently great, the vapour begins to liquefy; and the pressure then remains constant until the whole has become liquid. Further compression is comparatively a matter of great difficulty.

The whole process above described must take place with extreme slowness in order that the condition (§ 78) of constant temperature may be adhered to.

We may now suppose the temperature to be increased to, and maintained at, a definite value higher than that which it formerly had. If the pressure has the same value as it had at the commencement of the former process, the volume of the vapour will be greater than before, for all vapours expand when heated under constant pressure. And, if the pressure be increased as in the previous case, a precisely similar series of phenomena will be presented; the volume of the substance, however, being always larger than formerly under the same conditions as regards pressure. But one important difference will be noted—the change of volume during the process of liquefaction will be less than it was when the temperature was lower. Ultimately, when the temperature is sufficiently high, there is no sudden change of volume when the substance assumes the liquid condition. At still higher temperatures, the deviations from Boyle's Law become less and less marked.

At all temperatures above the limiting one at which sudden liquefaction ceases, the substance is called a *gas*; at lower temperatures it is termed a *vapour*.

With this explanation, no difficulty will be experienced in understanding the diagram on the opposite page. The volume, v, of a perfect gas is measured along the horizontal axis from a point not shown in the diagram. The scale is such that 1,000 times the reciprocal of the abscissa represents the pressure in atmospheres. The actual volumes, v', of carbonic acid gas are measured along the vertical axis. In this way a series of curves is shown which indicates the deviation of that gas from Boyle's Law at various temperatures. Portions of two of the nearly straight lines, which these curves would become if the gas were air, are drawn. The vertical portions of two of the curves indicate the stage during which liquefaction occurs—the almost horizontal parts belong to the liquid carbonic acid.

The substance is more or less compressible under given conditions of pressure and temperature than a perfect gas is, according as a line touching the curve (at the point satisfying these conditions) makes a greater or less angle with the horizontal axis than the angle whose tangent is v'/v times the tangent of the angle made by the corresponding perfect-gas line. Hence the substance, when in a condition resembling the liquid state, is less compressible than a perfect gas would be; and we thus see that hydrogen is less compressible than a perfect gas, because, under ordinary conditions of temperature and pressure, it is in a state more analogous to that of liquids than to that of gases. Were it examined under conditions similar to those of carbonic acid in the upper right-hand region of the diagram, *i.e.*, under sufficiently diminished temperature and pressure, it would almost certainly be found to have a smaller volume than Boyle's Law shows.

83. *Elasticity.*—All gases and vapours possess *perfect* elasticity of bulk. That is to say, they entirely recover their original bulk when allowed to do so by means of the removal of the distorting pressure. This may readily be proved by the simple experiment of inverting in water a glass vessel containing air or any gas which is not appreciably dissolved by the liquid. The gas may be subjected to, or relieved from, pressure by raising or lowering the glass vessel. The possibility of discharging a bullet from a gun, or of propelling a vessel or driving machinery by means of compressed gases furnishes another proof of their elasticity. And still another proof consists in the fact that they all convey sound, which would be impossible were they not elastic, just as it would be impossible if they were not capable of being compressed.

84. *Viscosity.*—We have already given a general definition of this term as the property in virtue of which there is resistance to shearing motion. But it is convenient to use the word as referring to a specific property (one independent of the size of the body, § 65). Hence we define viscosity as the tangential force per unit area of two indefinitely large parallel plane surfaces of the fluid which are at unit distance apart and move parallel to each other with unit relative speed. The tangential force per unit area of two such planes at a distance x apart, and moving with relative speed v, is $\tau v/x$, where τ is the viscosity.

In making an actual determination of the value of r in any gas, various forms of experiment based upon the above definition might be used. Clerk-Maxwell used a circular disc which hung in a horizontal plane from a vertical wire attached to its centre. By means of torsion of this wire the disc could be made to oscillate in its own plane. Two similar *fixed* discs were placed one on each side of the vibrating disc, and the gas occupied the intervening space. The disc would obviously oscillate more slowly in a viscous gas than in one which possessed small viscosity; and the quantity r may be determined from the results of such experiments.

This property varies very much from one gas to another. In hydrogen, carbonic acid, air, and oxygen, it increases from the first-mentioned to the last-mentioned, being about half as great in hydrogen as in air. It *increases* very markedly also with rise of temperature.

The subjoined table contains values of the viscosities of a few gases calculated from Graham's observations upon transpiration (§ 87), the units of length and time being respectively the centimetre and the second.

Hydrogen	0·000093
Carbonic acid	0·000160
Carbonic oxide	0·000184
Air	0·000192
Oxygen	0·000212

Tomlinson's recent determination of the constant for air gives the value 0·000177 at about 12° C.

The slow descent of clouds, or of fine suspended dust, in air is due to the viscosity of that gas. The weight of a drop of water, which causes its descent, is proportional to the cube of its diameter; but the resistance which results from viscosity is proportional only to the first power of the diameter. Hence, if the diameter of a drop be reduced to one-tenth of its original value, the weight becomes one-thousandth of what it was before, while the resistance is merely reduced to one-tenth of its previous amount. That is to say, the resistance is relatively one hundred times more effective than formerly.

85. *Diffusion.*—When two gases, which are not intimately mixed, occupy a certain volume, each gradually diffuses itself throughout the whole volume, so as to fill it just as it would have done had the other been absent. The only effect of the presence of the other gas (on the presumption that there is nothing of the

nature of chemical action between them) is that the time occupied in uniformly filling the space is greatly increased. The above process is called *diffusion*, and the corresponding property is *diffusivity*.

Experiment shows that the quantity of gas which passes in time t through an area a, perpendicular to which the rate of variation of density per unit of length is r, is proportional conjointly to t, a, and r. Hence, if q be this quantity, we have

$$q = \delta rta,$$

where δ is a constant (the diffusivity, or co-efficient of inter-diffusion) the magnitude of which depends upon the nature of the gases.

If r, t, and a are each unity, we get

$$q = \delta,$$

and so we define the diffusivity as *the quantity of the substance which passes per unit of time through unit area across which the rate of variation of density of the substance per unit length is unity.* The quantity r is generally called the 'concentration-gradient.'

The kinetic theory of gases leads to the conclusion that the co-efficient of interdiffusion of gases should be approximately in inverse proportion to the geometrical mean of the densities of the two gases under one atmosphere of pressure. The figures in the first column of the accompanying table give rough relative values of δ for pairs of gases. Relative values of the reciprocals of the geometrical means are given in the second column. The absolute values in square centimetres per second, determined by Loschmidt, are given in the third column.

Carbonic Acid and Air	1·0 ...	1·0 ...	0·1423
Carbonic Acid and Carbonic Oxide	1·0 ...	1·0 ...	0·1406
Carbonic Acid and Hydrogen ...	3·9 ...	3·8 ...	0·5614
Carbonic Oxide and Hydrogen	4·6 ...	4·8 ...	0·6422
Oxygen and Hydrogen	5·2 ...	4·5 ...	0·7214

The greatest deviation occurs with oxygen and hydrogen. This is probably due to molecular action between these gases.

86. *Effusion.*—The phenomena presented in the passage of gases through the pores of solids are of great interest, and have been elaborately investigated by Graham. The simplest results are obtained when the solid is practically of infinite thinness, is non-porous, and has a small hole drilled through it. If a gas is kept under constant pressure at one side of the solid, while a vacuum is

preserved at the other side, the process of passage of the gas is called *effusion*. The theoretical treatment of the question is extremely simple. The work done in the transference of unit volume of the gas is (§ 43) numerically equal to p, the pressure, for the total pressure on unit area then acts through unit distance. And the kinetic energy acquired is $\frac{1}{2}dv^2$, where d is the density or mass per unit volume, and v is the speed of the escaping gas. Hence the speed of escape, and therefore the quantity of the substance which passes through in one unit of time, is inversely as the square root of the density. The observed and calculated quantities for four substances are given in the subjoined table:

	Observed.	Calculated.
Carbonic Acid	0·835	0·809.
Oxygen	0·952	0·951.
Air	1·000	1·000.
Hydrogen	3·623	3·802.

It appears that hydrogen and carbonic acid pass through more rapidly and more slowly respectively than the above law would indicate. The reason for this will be seen in next section. The absolute values of the quantities which pass under atmospheric pressure, through an area equal to one-millionth part of a square centimetre, may be determined from the numbers given in the first column above, when the corresponding quantity is known for air. Now this can easily be calculated, for the pressure is that due to the height, h, of the homogeneous atmosphere (§ 58), and the square of the speed of escape is (§ 57) the product of $2g$ into this height. If a million of such holes be taken so that no one interferes with the action of another, the total area is one square centimetre, and the volume which escapes per second is $\sqrt{2gh}$. Hence, if the density of the air is 0·001293, the quantity which escapes per second is 50·9 grammes; so that the effusivities in centimetre-gramme-second units may be obtained from the numbers given in the table above by using 50·9 as a factor.

87. *Transpiration.*—When the non-porous septum, above referred to, is not thin, the small aperture becomes a tube of exceedingly fine bore, and the gas passes through by *transpiration*. Graham found that the rate of passage was altogether independent of the nature of the substance forming the walls of the tube. This suggests that a layer of the gas becomes deposited upon the interior of the tube, so that the gas has really to flow through a tube composed of its own substance in a highly condensed state. [It is well known indeed that most, and probably all, solids have a great power

of condensing gases on their surfaces or within their pores.] Hence we would expect that transpiration is a process which depends upon the viscosity of the gas. This is borne out by the fact that the rates of transpiration of oxygen, air, carbonic acid, and hydrogen are, in increasing magnitude, in the order in which these gases are here named, being fully twice as great in hydrogen as in air—an order which is the exact reverse of their order as regards viscosity. Hence we conclude that the abnormality in the rates of effusion of hydrogen and carbonic acid are due to viscosity, the hole in the thin plate acting to some extent as a short tube.

88. When the pores of a substance through which a gas passes are extremely narrow (as in fine unglazed earthenware), and the substance does not condense the gas in its pores, the rate of passage follows the ordinary law of diffusion or effusion, *i.e.*, gases pass through at rates which are inversely as the square roots of their densities. Hence we have a means—known as the method of *Atmolysis*—by which to separate a mixture of two gases of different densities. If the mixture be placed inside a porous earthenware vessel, the less dense gas passes through most quickly, so that, when the process has gone on for some time, we have two portions of gas, one containing in most part the less dense gas, the other composed mostly of the denser one. The process may be repeated so as to separate the two constituents to any desired extent.

It has been already mentioned (§ 62) that carbonic oxide passes rapidly through red-hot iron; and hydrogen passes through palladium, and even platinum, at ordinary temperatures.

In some cases the gas combines chemically with the substance on one side, diffuses through it, and is given off on the other side. This occurs with indiarubber.

CHAPTER VIII.

PROPERTIES OF LIQUIDS.

89. *Compressibility.*—Liquids, like gases, convey sound, and are therefore compressible and elastic. But a liquid differs from a gas, in that its compressibility is usually extremely small. It differs, also, as widely in respect of the law of compression. An inspection of the diagram of § 252 will show that a vapour such as carbonic acid becomes more and more compressible as it approaches the liquefying stage, while, during liquefaction, the compressibility is infinite. For example, the right-hand portion of the isothermal of $21^\circ{\cdot}5$ is practically a straight line, and therefore the rate at which the volume increases with pressure is constant. But the compressibility is equal to this rate′ divided by the volume, and the volume diminishes as the liquefying stage is approached, so that the compressibility increases. The change is in the opposite direction when the whole substance has become liquid; the compressibility is extremely small, and diminishes as the pressure increases.

The earlier determinations of the compressibility of liquids were made by means of an apparatus called the *piezometer*, and the more perfect modern appliances all work on the same principle. This apparatus consists of a large glass bulb, having a narrow carefully graduated stem, which is open at the top. The internal volumes of the bulb and of the stem are accurately measured. The liquid, whose compressibility is to be determined, fills the bulb and part of the stem. A small column of mercury suffices to separate the liquid inside the bulb from water which fills a strong glass vessel, inside of which the bulb is placed. The outer vessel is closed, and pressure is applied by screwing in a plug so as to diminish the internal volume.

The pressure is communicated to the liquid inside the bulb, since there is a complete liquid connection through the stem; and its amount may be measured by means of the compression of air contained in a glass tube, which is closed at the upper end, and is also placed inside the outer vessel.

If the glass were incompressible, the compression of the liquid would be at once found by means of the extent to which the mercury index descended in the stem. If the liquid were incompressible, while the glass was capable of compression, the index would rise. If the liquid and the glass were equally compressible, the position of the mercury would not alter. Hence we see that this experiment really gives the *difference* between the compressions of the glass and the liquid, so that we must first know by experiment the compressibility of the glass of which the bulb is formed. This, as we shall see in the next chapter, is a comparatively simple matter.

In the second column below, the numbers represent compression per atmosphere at 0°C. The numbers in the third column give the compression per atmosphere at temperatures Centigrade which are recorded in the last column.

Mercury	0·00000295	0·00000187	15·0
Olive-oil	0·00004859	0·00006174	18·3
Water	0·00005030	0·00004563	22·9
Bisulphide of carbon	0·00005392	0·00006378	17·0
Alcohol	0·00008282	0·00009595	17·5
Ether	0·00011560	0·00014770	21·4

Water reaches a minimum compressibility at a temperature of about 63°C.

The compressibility of all liquids is lessened by increase of pressure.

90. *Elasticity.*—All liquids, like gases, possess perfect elasticity of bulk, and, in common also with gases, have no elasticity of form.

91. *Viscosity and Viscidity.*—Viscosity is very much more marked in liquids than in gases, and varies greatly from one liquid to another. The slowness of the descent of fine mud in water is due to the viscosity of that liquid, and the slowness of the fall of fine rain-drops is caused by the viscosity of air. Glycerine is one example of an extremely viscous liquid, while, in comparison with it, sulphuric ether has little viscosity.

One very simple method of determining the viscosity of a liquid consists in forcing it under pressure through a cylindrical tube of very fine bore. The quantity which passes through per unit time is directly proportional to the difference of pressure per unit length of the tube and to the fourth power of the radius, and is inversely proportional to the co-efficient of viscosity.

Viscosity diminishes rapidly with increase of temperature.

In the second column of the table below are given values of viscosity, in centimetre-gramme-second units, observed at the temperatures (Centigrade) recorded in the third column.

Water	0·0178	0
,,	0·0131	10
,,	0·0102	20
,,	0·0056	50
,,	0·0036	80
Glycerine	42·18	2·8
,,	4·94	26·5

Viscidity is a related property in virtue of which a liquid can be drawn out into long threads. Other things being equal, a liquid is viscid in proportion to its viscosity; but the molecular forces produce another effect besides viscosity, which acts so as to prevent viscidity (§ 101).

92. *Diffusion.*—Under the diffusion of liquids we include the diffusion of solutions of solids.

Diffusion of liquids is a very much slower process than diffusion of gases. If a solution of bichromate of potassium be carefully introduced at the foot of a vessel containing water, the process of interdiffusion may go on for months before appreciable uniformity is attained.

Many methods (electrical, optical, etc.) exist, by means of which the co-efficient of interdiffusion (the definition of which is identical with that given in § 86) may be determined.

In one method, due to Graham, communication is established between two vessels, each of which contains a liquid capable of diffusing into that contained by the other. Special care is taken to avoid the production of currents whether in the act of establishing communication or because of difference of density of the liquids. The communication is closed after a definite time, and the extent to which diffusion has gone on is determined. A series of precisely similar experiments is made, each experiment of the series lasting for a different interval of time, and the diffusivity is determined from the results.

The numbers (deduced from Graham's results) in the second column below give, when divided by 10^6, diffusivities, with reference to water, in centimetre-second units—the temperature being about 10°C. Increase of temperature causes increase of diffusivity.

Common salt	1·050
Sugar	0·360
Caramel	0·058

93. *Osmose.—Dialysis.*—Diffusion of liquids can take place through animal membrane, such as a piece of bladder. The less dense liquid passes through most quickly. If a vessel containing a strong solution of sugar be closed tightly by means of a membranous substance, and then be immersed in a vessel of water, the contents of the inner vessel will rapidly increase, and may finally cause the membrane to break. The process of such transference is called *osmose.*

Liquids may be broadly divided into two classes with reference to the readiness with which they pass through animal membranes. Crystalloid substances, such as common salt, sugar, etc., pass easily through when in solution; but solutions of colloid substances, such as glue, can scarcely pass at all. This is the basis of the process of *dialysis*, which is used for the separation of a mixture of colloid and crystalloid bodies. The mixture is separated from pure water by a portion of animal membrane, through which the substances pass in very disproportionate quantities. One or two repetitions of the process are sufficient to practically separate the two constituents of the mixture.

The method is essentially analogous to the method of atmolysis which was described in § 88.

94. *Cohesion.*—Cohesion is that property which, apart from the mere gravitation of the parts as a whole, results in the clinging together of portions of matter whether of the same or of unlike kinds. It may be regarded as the result of the so-called molecular forces. (See, again, § 67.) When a body has been pounded down, so that its parts have been separated beyond the range of the molecular forces, cohesion may be brought about again by the application of pressure sufficient to place the molecules once more within the range of their mutual forces. In the case of a liquid, it is sufficient to merely place the separated parts in contact. (For further treatment, see under Properties of Solids.)

95. *Capillarity.*—It is a well-known law of hydrostatics (§ 58) that the pressure has the same value at all points of a fluid which are at the same level, so that we should expect that the level must be the same at all surfaces of a continuous fluid mass which are exposed to the atmosphere. Yet, in some cases, this is far from being the fact.

If a fine capillary tube be inserted in some liquids, the level is higher inside the tube than it is outside, while in other liquids the reverse is the case (Fig. 43). Thus water rises inside a glass tube while mercury descends.

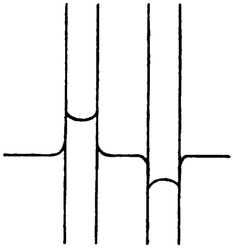


FIG. 43.

These phenomena (called *capillary* phenomena) seem to be in direct violation of the above-mentioned law of hydrostatics, but in reality they are in strict accordance with it.

96. In proof of this we observe, first, that the surface of a liquid which *rises* in a capillary is always *concave* upwards, while the surface of one which *descends* is invariably *convex* upwards.

Next, we observe that, if a surface, originally plane, is under tension and is curved, there must be more pressure on the concave than on the convex side, otherwise the surface would once more become plane, because of its tendency to shrink. Hence, if we can show that there is tension in the surface films of liquids, it follows that there is more pressure on the concave side than on the convex side of the curved surface of a liquid.

But it is well known that the surfaces of liquids tend to become as small as possible. Many examples of this fact may be brought forward. A soap-bubble contracts of itself if the air inside it be in communication with that outside; and the mere fact that the soap-bubble is naturally spherical constitutes another proof, for the sphere is the minimum surface which can enclose a given volume. For the same reason rain-drops are spherical—which is proved by the perfect circularity and definiteness of the rainbow. Again, if some alcohol be dropped on a thin layer of ink, the surface of the ink will decrease, while that of the alcohol is increased, because of the excess of the surface-tension of ink over that of alcohol.

Let us suppose now that the surface of a liquid in a narrow tube becomes hollow upwards. Just underneath the outside plane surface there is atmospheric pressure, while just below the curved surface the pressure is less. At some point, p (Fig. 44), lower down in the tube, the pressure (which is there increased because of the weight of the liquid) is equal to that of the atmosphere. Hence, *in strict accordance with hydrostatic laws*, the liquid *must* rise until the point p is at the level of the surface of the liquid outside.

Similar considerations explain the depression, in a narrow tube, of a surface which is convex upwards.

It only remains to explain *why* the surface becomes curved. We shall assume that the tube is of glass, that the liquid is water, and that the surrounding atmosphere is air. The surface of the liquid in contact with the glass is under tension, the amount of which per unit breadth of the film we may denote by ${}_wT_g$. There is also a film of condensed gas on the surface of the glass, the tension of which per unit breadth we may similarly represent by ${}_aT_g$. The particles of the liquid at the edge will therefore be pulled upwards or downwards, according as ${}_aT_g$ is greater than, or less than, ${}_wT_g$. In the case assumed, ${}_aT_g$ is greater than ${}_wT_g$, and so the water surface becomes concave upwards. The tension, ${}_aT_w$, of the water surface which is in contact with the air—which formerly acted straight outwards from the walls of the tube—now acts downwards at an angle a with the side of the tube (Fig. 45). The inclination tends to be such that the tension ${}_wT_g$, together with the downward component (§ 28) of ${}_aT_w$, just balances the upward tension ${}_aT_g$.

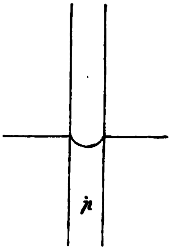


FIG. 44.

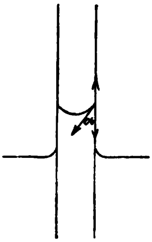


FIG. 45.

97. The surface of water in a narrow glass tube is hemispherical, for the angle a is at least practically zero, and probably vanishes entirely when care is taken to have the water free from all dissolved impurities. It is very essential to keep the surface of the liquid free from impurity, and to ensure that the surface of the solid is chemically clean. The slightest trace of grease might entirely prevent the liquid from rising.

The angle a, which is called the *angle of contact*, may be found experimentally by the following method. Let AB (Fig. 46) be a plane plate of the glass (or other solid), and let it be dipped into the liquid CD. If the liquid rises and wets the solid, making an acute angle a with it, it is evident that when AB is inclined at the angle a to CD, the level of the liquid is unaltered at that side of the plate which faces upwards. We may now find a by direct measurement. The figure, if turned upside down, corresponds to the case of a liquid which descends in a capillary tube. Magie finds that the angle is zero, with ethyl alcohol, methyl alcohol, chloroform, formic acid, and benzol, in glass tubes; in the case of mercury and glass its value is about 142°.

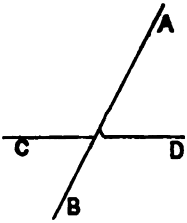


FIG. 46.

98. We shall now consider the height to which a liquid will rise in a tube of given bore. Let r be the radius of the tube, and let T be the vertical component of the tension per unit breadth of the surface separating the liquid from the air. Since T is the difference of the upward and downward tensions per unit breadth, the quantity Tc represents the total upward pull if c is the internal circumference of the tube. This is balanced by the weight of the raised liquid, which, in a given tube, is proportional to the difference of level inside and outside. Since the volume occupied by the raised liquid is proportional to the square of the radius of the tube, the height will, in the case of a given liquid, be inversely proportional to the radius.

When the liquid rises between two parallel plates, whose distance apart is equal to the diameter of the tube in the preceding case, we may consider a column of the liquid whose breadth is also equal to the diameter. This square column is only pulled up on two sides (*i.e.*, on half) of the perimeter of its upper face. Hence, all other conditions being as before, the liquid can only rise half as high as it does in the tube of equal diameter.

The value of the surface tension may be found by observations on the height to which a liquid rises in a capillary tube, and by various other methods. Quincke gives approximately the following values in grammes weight per linear centimetre at a temperature of about 20° C.:

Mercury	0·550
Water	0·075
Bisulphide of carbon	0·033
Alcohol	0·026

The height to which the liquid rises is inversely proportional to *d*, the distance between the plates. And since, if the plates be not parallel, but be placed in contact along one of their vertical edges, *d* will be proportional to the distance from the common edge, the liquid, rising highest where *d* is smallest, will meet each plate in a curve which is a rectangular hyperbola. The axes of the hyperbolas will coincide respectively with the common edge, and with the lines in which the level surface of the liquid meets the plates.

99. The results of § 96 enable us to explain the strong 'attraction' of two parallel glass plates between which a drop of water is placed. For, since the water becomes concave outwards, the pressure inside it is less than that of the atmosphere, and hence the plates are *pressed*, not attracted, together. The lifting of a stone by means of a leather 'sucker' is similarly explained.

The plates would be apparently repelled apart if the liquid did not wet them, *i.e.*, if it became convex outwards. And it is easy to prove that this would result also if only one plate were wet. The liquid would rise to a greater height outside the one plate, and would descend farther outside the other, than it would rise or fall in the interior space. At the part *ab* (Fig. 47) there is less than atmospheric pressure outside (for the liquid is concave upwards), while inside there is atmospheric pressure. Also, at *cd*, there is more pressure on the inner, than on the outer, side. Thus, from both causes, the plates are pressed apart.

100. In the proof of § 56 we may suppose that, instead of a cord stretched in a circular tube, we have a film of unit breadth stretched over a cylinder. The pressure per unit area will therefore be T/R, T being the tension exerted across unit breadth.

If the film be stretched over a spherical surface of the same radius R, the pressure would have the value 2 T/R, for there are now equal curvatures in two directions at right angles to each other. [The investigation of § 98 furnishes a special proof of this. For if the angle of contact is zero, the surface of the liquid is hemispherical in the circular tube, and is cylindrical in the space between the parallel plates. But we have seen that, if the radii of the cylinder and the sphere are equal, the liquid rises twice as high in the tube as it does between the plates. In other words, the pressure towards the centre of curvature of the film, which supports the weight of the elevated liquid, is twice as great in the first case as in the second.]

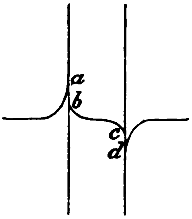


FIG. 47.

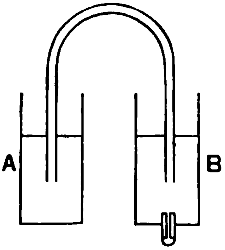


FIG. 48.

In a soap-bubble, we must remember that the liquid film has *two* surfaces; so that, when the bubble is spherical, the air in the interior is subjected to a pressure which is greater than that on the outside by the quantity 4T/R.

The above considerations indicate a method, due to Lord Kelvin, of measuring the value of T. A capillary tube is inserted in the bottom of a vessel B (Fig. 48) which is partly filled with the given liquid, and is connected by a syphon with a vessel A which also contained the given liquid. By raising or lowering A, the level of the liquid in B may be altered at pleasure. The liquid will pass through the capillary tube, and will gather into a drop at the lower end of it; but this drop will not fall away unless the difference of level, h, between its lowest point and the free surface of the liquid in B is too great. The inward pressure per unit surface of the drop is $p+2T/r$, where p is the atmospheric pressure and r is the radius of the drop (measured by micrometric methods). The outward pressure per unit surface, due to the weight of the liquid and the pressure of the atmosphere, is (§ 58) $p+hs$, where s is the specific gravity of the liquid. Hence $2T=rhs$.

101. We have hitherto regarded surface-tension as an observed fact merely, but it is easy to see that it is a necessary result of the mutual potential energy of molecules, or, as we may put it, of the molecular forces. Let p' (Fig. 49) be a molecule in the liquid,

situated at a greater distance from its surface than the range through which the molecular forces are sensible. Draw a sphere from p' as centre with the range of the molecular forces as radius. There is no mutual action between p' and any molecule outside this sphere; and it is equally attracted on all sides by the molecules inside the sphere. But any particle p, which is nearer the surface of the liquid than the given distance, is pulled inwards on the whole by the molecular attraction of the interior particles. And this inward pull on the surface particles produces the same effect as, and will obviously be manifested as, a surface-tension tending to diminish the external periphery of the liquid.

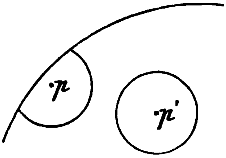


FIG. 49.

102. The tension of a sheet of indiarubber increases in proportion to the augmentation of the surface, but the tension of a liquid film remains absolutely constant (at least, through extremely wide limits) when the area of the surface is altered. If left to itself, the indiarubber will contract until the area of its surface once more attains its original value; but the liquid will contract until its surface becomes as small as possible.

Consider a film of breadth b, the tension of which per unit breadth is T, so that the total tension is Tb in the direction of the length of the film. If the length of the film be increased by the amount l, the work done in the process is Tbl (§ 43). But this is equal to TS, where S is the increase of surface. Thus the work is directly proportional to the increase of area, and we may look upon T as the amount of work done per unit increase of area instead of a tension per unit breadth.

Taking this fact in conjunction with the result of last section, we can now obtain an expression for the exhaustion of potential energy of molecular separation (the work done by the molecular forces) when two separate masses of the same liquid are placed in contact over a given area. Let S be the area, so that 2S is the diminution of surface of the two masses. The work done is 2TS.

Let T and T′ be the surface-tensions of two different liquids, and let t be the tension of the surface separating the two when placed in contact. The work done by the molecular forces when they come into contact is obviously $(T+T'-t)$ S. The above result is a particular case of this, for, when the two liquids are identical, we have $t=0$ and $T=T'$.

If in any case the work done is greater than $(T+T')S$, *i.e.*, if t be negative, the surface of separation of the two liquids must increase. This it may do by becoming puckered; and, the smaller the scale of the puckering, the greater will be the increase of surface. Kelvin regards this invisible replication of the separating surface as the commencement of the process of diffusion.

103. The surface-tension of a liquid diminishes rapidly with rise of temperature, and it vanishes entirely at the critical temperature (§ 20).

The saturation pressure (§ 249) of the vapour of a liquid depends upon the temperature. But the temperature being fixed, it also varies with the curvature of the surface of the liquid, being greater the more convex outwards the surface is. Hence small drops of water in a cloud evaporate, the vapour being deposited upon the larger ones.

From the fact that the surface-tension of liquids decreases with rise of temperature, we might deduce, by the principle of stable equilibrium (§ 15), the result that sudden extension of a film will produce a fall in its temperature. For, since the system is in stable equilibrium, it follows that extension of the film will produce an effect which results in an increase of the force resisting the extension. It will, therefore, cause diminution of temperature. This is known to be the case.

The principle of dynamical similarity (§ 49) shows at once that the square of the fundamental period of vibration of a (weightless) liquid sphere is directly proportional to the density of the liquid and to the cube of the radius of the sphere, and is inversely proportional to the surface-tension. It also shows that the period of fundamental vibration of a (weightless) soap-bubble is independent of its linear dimensions.

CHAPTER IX.

PROPERTIES OF SOLIDS.

104. *Compressibility and Rigidity.*—The compressibility of a solid is defined in precisely the same way as that in which we have already defined the compressibility of a liquid or gas. It is the ratio of the fractional change of volume to the change of pressure which produces it. The reciprocal of this quantity is called the resistance to compression, and is usually denoted by the letter k.

The compressibility is most readily determined by measurement of the alteration of length of a rod of the substance to which known hydrostatic pressure is applied. If p' be the fractional alteration of length, the fractional alteration of bulk is approximately $p=3p'$. For l, b, and t, representing respectively the length, breadth, and thickness of a rod of the given substance, the new volume is $lbt(1-p')^3$, which approximately is $lbt(1-3p')=lbt(1-p)$. In all actual cases the value of p' is so small that any power higher than the first may be neglected. [It must be observed that we assume the substance to be *isotropic*, *i.e.*, its properties are independent of direction. If this were not so, p' might have different values in different directions. This assumption will be adhered to throughout the chapter.]

The rigidity of a solid is the measure of its resistance to change of shape. Let ABCD (Fig. 50) be a cube of a given solid, the edges of which are of unit length. Let equal tangential forces, of magni-

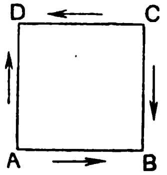


FIG. 50.

tude T per unit area, be applied to the opposite faces AB and CD, and let them act in the directions indicated by the arrows. These forces will produce shearing of the cube, but they will also produce rotation in a direction opposite to that of the hands of a watch. To prevent this rotation, tangential forces equal to the former may be applied to the opposite faces AD and CB. The result of the application of this set of forces is that the square section shown in the figure becomes rhomboidal, the angles at B and D being made less than a right angle by the same amount that the angles at A and C are made larger than a right angle. If θ be the change of angle, the rigidity, which is usually denoted by the letter n, is given by the formula

$$n = T/\theta.$$

105. Let us denote the three pairs of parallel faces of a unit cube by the letters A, B, and C. Similarly we shall speak of the edges joining the A faces as the A edges, and so on.

Let unit normal pressure per unit area be uniformly applied to the A faces. This will diminish the A edges by an amount l, and will increase the B and C edges by a common amount l'. Now let unit normal tension per unit area be applied to the B faces. This will increase the B edges by the amount l, and diminish the C and A edges by the amount l', small quantities of the second order of magnitude being neglected. The result is that the A edges and the B edges are respectively diminished and increased by the amount $l+l'$, while the C edges are unaltered in length. Hence there is no alteration of volume.

Now this result might have been produced by the method of last section. If any point in the face DC (Fig. 51) of the unit cube be slid forward relatively to a point in AB through the (very small) distance s, the increase of length of the diagonal DB is the resolved part of s in the direction of the diagonal. It is, therefore, equal to $s/\sqrt{2}$. Similarly the decrease of length of AC is $s/\sqrt{2}$. The given tangential forces are obviously equivalent to a pressure-stress (§ 50)

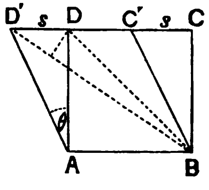


FIG. 51.

parallel to AC of magnitude T$\sqrt{2}$ per area $\sqrt{2}$, *i.e.*, T per unit area, together with a tension-stress parallel to DB of the same magnitude. If, now, we let T$=1$, $s/\sqrt{2}$ is the alteration of length of the diagonal which contains $\sqrt{2}$ units, and so $s/2$ is the fractional change of length in the direction of the diagonals due to unit tension parallel to BD and unit pressure parallel to AC. Hence, equating the results of the two methods, we get $s=2(l+l')$. But $s=\theta$, the change of angle of the unit cube. Hence

$$l+l=\frac{1}{2n} \quad \ldots\ldots \quad (1).$$

106. Unit pressure per unit area on the A faces shortens the A edges by the amount l, and increases the B and C edges by the common amount l'. The quantities l and l' being extremely small, if unit pressure be now applied to the B and C faces, all the edges of the cube will be diminished by the amount $l-2l'=p'$. Hence

$$l-2l'=p/3=1/3k \quad \ldots\ldots \quad (2).$$

The quantity $1/l$ (the reciprocal of the fractional change of length of a rod under unit tension or pressure per unit of its transverse sectional area) is called *Young's Modulus.* From (1) and (2) we find

$$l=\frac{1}{9k}+\frac{1}{3n}=\frac{3k+n}{9kn}. \quad \ldots\ldots \quad (3).$$

This formula enables us to determine the value of either l, k, or n, provided that we know the values of the other two. The following table gives the values of *Young's Modulus* for a few substances in grammes weight per square centimetre. To reduce to dynes per square centimetre, the numbers must be multiplied by 981, the value of gravity in centimetre-second units:

Steel (cast)	$1955(10)^6$
Steel (wire)	$1881(10)^6$
Iron (cast)	$1375(10)^6$
Iron (wire)	$1861(10)^6$
Copper (wire)	$1185(10)^6$
Zinc (drawn)	$873(10)^6$
Lead	$177(10)^6$
Ash	$113(10)^6$
Oak	$103(10)^6$
Glass (flint)	$615(10)^6$

107. The rigidity n is not found in practice directly by the process

which was described in § 65. It may be found by determining the moment of the couple which is required to twist a circular cylindrical rod of the substance through a given angle about its axis. This is easily seen by means of elementary considerations. Let the rod be of unit length, and imagine it to be divided into a very large number of very small similar and equal rods by means of planes parallel to its length. Let r be the distance of one of these rods from the axis, which we may assume to be vertical. The amount by which, with a given twist θ of the whole rod, the upper end of this little rod slides forward relatively to the lower end, is proportional to r and to θ. So long as Hooke's Law (§ 111) holds, it is also proportional to the tangential force which is applied to the end of the rod to produce the twist. Hence the moment of this force (§ 53) is proportional to $r^2\theta$. But the number of such little rods per similar and similarly situated area of the end of two such cylinders is proportional to the (similar) areas. Hence, if r now represents the radius of the cylinder, we see that the moment of the total couple is proportional to $r^4\theta$. Let c be the couple, while n is the rigidity, and k is a constant, and we get

$$c = kn\theta r^4.$$

A strict investigation shows that the value of k is $\pi/2$.

Even a simpler experimental method consists in attaching to one end of the rod a body whose moment of inertia about the axis of the rod is very great. If the rod be firmly clamped in a vertical position by its upper end, the body being attached to its lower end, the time of oscillation of the whole system about the axis depends upon the value of n.

108. The following table gives, in the same units, the values of n for some of the substances for which Young's Modulus was given in § 106. From these numbers the values of k may be calculated by means of (3) § 106. The observed values of k are given in the last column. There is considerable discrepancy between the calculated and the observed values, but this need not produce surprise, for there are many causes of variation in the experimental results. The value of r in the expression for the rigidity is usually small, so that a large error may occur in its measurement; and—even if r were uniform throughout the rod, which is rarely the case—four times this error will be produced in the calculated values of n, since r is involved to the fourth power. Again, the substances may not be really isotropic; and the special physical treatment—*e.g.*, the drawing out of a wire—to which an originally isotropic substance is subjected will frequently make it non-isotropic.

	n		k
Iron (cast) ...	$542(10)^6$	...	$938(10)^6$.
Copper (wire)	$445(10)^6$	...	$1172(10)^6$.
Zinc (drawn) ...	$360(10)^6$	...	$506(10)^6$.
Glass (flint) ...	$243(10)^6$	...	$437(10)^6$.

109. The *flexural rigidity* of a bar in a given plane is measured by the moment of the couple which is required to produce unit curvature in that plane. In similar and equal bars of different substances it is directly proportional to Young's Modulus; and, in different bars of the same substance and of similar though unequal section, it is proportional to the square of the sectional area. However a bar be bent, the locus of the centres of inertia (§ 52) of all the transverse slices is unaltered in length. The length of all other lines is either increased or diminished. This shows why Young's Modulus is involved.

Consider a rectangular rod, of thickness d and of breadth b. Let it be bent uniformly to unit curvature, and let us suppose that b and d are small in comparison with the radius of curvature. If we further imagine the rod to be composed of a very large number of rods, whose cross-sections are similar to that of the given rod, and whose lengths are equal to the length of the given rod, it is easy to see from similarity that the couple which is required to produce the curvature must be proportional to b, the breadth of the rod measured perpendicular to the plane of bending. Also the elementary rods which are further away from the centre of curvature than the central plane of the given rod are extended in proportion to their distance from that plane, while those nearer to the centre of curvature are shortened in the same proportion. Hence we see that the total force must be proportional (so far as this effect goes) to d and to m—Young's Modulus; and therefore the moment of the couple must be proportional to m and to d^2. But the number of little rods, of a given size, is proportional to d; and therefore, finally, the moment c must be proportional to b, to m and to d^3—say

$$c = fmbd^3,$$

where f is a constant. Thus we see that the flexural rigidity of a rectangular bar is proportional to its breadth and to the cube of its thickness in the plane of bending.

110. *Elasticity.*—All solids possess elasticity, both of form and of bulk, to a greater or less extent. Within limits (which vary greatly in different substances) the elasticity is perfect, *i.e.*, the original form or volume is entirely regained; but if too great stress be applied,

the removal of the stress will not at once be—often will never be—followed by complete recovery from distortion. The limits (zero on the one hand, and a certain maximum on the other) of stress or strain, between which the elasticity is perfect, depend not only on the chemical nature of the substance, but also, as will be seen below, on the mechanical treatment to which the particular specimen under investigation has been subjected, and on various other conditions. Steel is a good example of the class of substances which have the limits wide apart, and which may break before much permanent distortion is apparent. On the other hand, lead can scarcely, under any conditions, recover entirely from a distortion, however slight. It exemplifies the class of substances whose limits lie close together, and which may suffer large permanent distortion before they break.

An elastic solid, if it be kept distorted for a considerable time, and then be released so slowly that it does not vibrate, will not in general at once assume its final form, but will gradually creep towards it. If it be consecutively distorted, first for a considerable time in one sense, and then for an equal time in the opposite sense, it will slowly recover from the second distortion to some extent, and will then undo the quasi-permanent part of the first.

That strain, or part of a strain, which disappears wholly on the removal of the distorting stress, is called *temporary strain;* that which is observed after the complete removal of the distorting stress is called *set.* The latter may, and usually does, diminish as the time which has elapsed since the complete removal of distorting stress increases. The rate of diminution becomes less and less, and finally the set reaches a fixed value, which is called the *permanent set.* This gradual diminution, together with the converse effect of the gradual increase of set under continued constant stress, is called *after-action* by the Germans.

In 1835 Weber investigated the laws of permanent set of a stretched fibre. In 1837, Hodgkinson, as the result of experiments on cast-iron, came to the conclusion that ‘the maxim of loading bodies within the elastic limit has no foundation in nature’; that is, permanent set is produced by any stress, however small. In 1842 and 1848 Wertheim observed that permanent set occurred *in a previously undistorted body* as the result of any stress, however small; and in 1848 this limitation to Hodgkinson's statement was independently pointed out, as a deduction from theory, by Professor James Thomson. On the other hand, a body previously distorted in a given sense may be again distorted to a smaller, or the same, extent in the same sense without the production of a new permanent set.

Between 1858 and 1862 G. Wiedemann made statical experiments on the torsion of rods, in the course of which he verified Wertheim's observation; and in 1880 he published the results of more extended experiments of the same kind. One of these results is that, after repeated twistings, alternately in opposite directions, by a given couple, the set of a rod becomes constant; and if the rod be again twisted, by increasing couples, in the direction of the last twist, the strain (measured from the position of set) is practically proportional to the stress so long as the original value of the couple is not exceeded. Another is that the reversed couple produces a greater strain, measured from the last set, than does the equal direct couple; the difference in this case corresponds to the change of zero produced by the reversal of the couple—that is, to the new set. Again, by repeated reversals of twist under a given couple, the total torsion and the set diminish to fixed minimum values. Also, in the case of torsion in one direction, the values of the total torsion and of the set increase at increasing rates as the couple increases, and the latter relatively at a greater rate than the former: these values for a given couple increase to maxima by repeated applications of the couple, and this increase is also relatively greater in the set than in the total torsion. Wiedemann remarks that the approach of the position of final set to that of final total torsion in this case is a phenomenon of the same kind as the narrowing of the limits of total torsion and of set by repeated reversals of a given couple; the only difference is that the negative couple is zero. He calls the process by which the wire is brought into the steady state as regards total torsion and set the process of *accommodation*.

In 1865 Lord Kelvin showed that the elasticity of a wire is greatly diminished if the wire be kept oscillating torsionally for a long period of time. That this is so is made evident by the fact that the oscillations then die away much more rapidly when the wire is left to itself after being set in oscillation. The elasticity is said to be '*fatigued*' by the process.

111. We may conveniently define elasticity as that property in virtue of which stress is required to maintain strain. Within the limits of elasticity, the necessary stress is proportional to the strain in all cases of distortion of a solid as well as in the special cases given above. This is known as *Hooke's Law*, and is usually stated in the form '*Distortion is proportional to the distorting force.*'

The constancy of the pitch of a note which is given out by a musical instrument proves that Hooke's Law is obeyed by the vibrating substance within the given limits of distortion. For, if distortion is proportional to distorting force, acceleration is proportional to displacement, and so the time of vibration must be constant (§ 38).

112. *Viscosity.*—Viscosity, or internal friction, is apparent in solids as well as in fluids. The vibrations of a tuning-fork die away at a greater rate than can be accounted for by the communication of energy to the air in the production of sound. Some process which we may call internal friction transforms the original energy partly into the form of heat in the material of the instrument. The phenomenon of fatigue of elasticity is also due to this cause, the internal friction being increased by the process which induces 'fatigue.'

Lord Kelvin showed that, when the range of torsional oscillation of a wire is 'well within the palpable limits of elasticity,' the decrease of the range of oscillation in a given number of oscillations is proportional to the range. This result was subsequently confirmed by Tomlinson. When the palpable limits of elasticity are exceeded, the range decreases with a rapidity much greater than that indicated by this law. Peddie has investigated the law of decrease of the range when the oscillations are large enough to produce marked set, and he has shown that it may be deduced from the assumption that the loss of energy per oscillation is proportional to a power of the range, the numerical value of the power increasing with the initial range. He has shown also that Kelvin's law for small oscillations follows, if the power becomes the second power when the initial range is small; and he has deduced an expression for permanent set, in terms of total torsion, which agrees well with Wiedemann's observed values. The value of the power is increased by 'fatigue.'

Lord Kelvin pointed out that the decay of the oscillations might be due to true viscous resistance, *i.e.*, resistance to change of shape depending on the rate of change, in which case the elasticity is perfect within the limits of the experiment. On the other hand, he remarked that, if it were due to dependence of the elastic resilient force on previous conditions of strain, the phenomenon would be continuous with imperfectness of elasticity, and this was indicated to some extent by the phenomenon of 'fatigue.'

His final experiments (alluded to above) were made with small distortions, within the palpable limits of elasticity, to test the former theory. The loss of energy per vibration was greater than could be accounted for by dissipation ensuing on change of shape of an elastic solid, and it was also immensely greater than that due to resistance by the air. He considered also that the results showed a loss of energy much greater than any that could be accounted for by imperfection of elasticity.

Though Kelvin's law was precisely that which would arise from viscous resistance, yet, were this the cause, decrease of the range per oscillation should vary inversely as the period; but this was far from being so. Tomlinson finds that the decrement is nearly independent of the period in the more 'non-viscous' metals, and that it increases markedly with the period in other metals. Kelvin remarked that the result was such as might be due to 'after-action' or imperfection of elasticity.

Wiedemann's statical observations showed that a wire which had once been twisted behaved in quite different ways according as it was twisted to one side or the other of its existing position of equilibrium; whence he concluded that 'the hypothesis according to which the decay of torsional oscillations is due to internal friction depending on the velocity alone, can no longer be maintained.' Nor could the elastic after-action alone account for it. He then stated his own theory as follows: Let the wire be 'accommodated' by frequent rotations to and fro [see § 110], and therein at last be temporarily twisted in the positive direction, describing an angle $+a$, while the molecules may be rotated so that the lower ends of their axes, looked at from the axis of the wire, describe an angle $+\alpha$, say to the left. The direction of this rotation follows from my magnetic experiments. If the wire be slowly brought back into the permanent torsion-position $+b$, the axes of the molecules will retain a portion $+\beta$ of their rotation to the left. If the wire now receives an impulse in the positive direction, which again elongates it up to $+a$, according to the laws of perfect elasticity it will swing back again to the position $+b$. If it now swings beyond this position farther to the right, and if the molecules in their rotation had no friction on one another at all to overcome, it would arrive at the elongation $-a$, since the same force that twists it from $+b$ to $+a$ twists it in the contrary direction from $+b$ to $-a$, while the axes of the molecules would be rotated just as far $(-a)$ to the right as previously to the left. Again, with perfect elasticity the wire would go back to the position $-b$, in which the molecules would be rotated $-\beta$, and so forth. The to-and-fro motions of the wire, between $\pm a$ and $\pm b$, are perfectly elastic; therefore, the performances of work in the swingings outwards and the swingings back again must, within these limits, completely compensate one another. In fact, however, there results a diminution of the amplitudes of oscillation; hence the loss of *vis viva* therein can only correspond to the work which is expended for the alteration of the positions of equilibrium, or the rotation of the molecules from $+\beta$ to $-\beta$, which determine it.

The kinetic theory of the viscosity of gases, as developed by Maxwell, asserts that viscosity is due to interchange of momentum between relatively-moving portions of the substances—this interchange being effected by the passage of molecules from one portion to the other. In the same way the viscosity of liquids is explained. There is essentially a passage of molecules from one group to another. Such passage does not take place in a perfectly elastic solid; but there may still be interchange of momentum in the relative motion of the constituents of a group, and therefore true viscosity in a solid. Yet, if the potential energy of deformation of a group is large in comparison with the kinetic energy of average relative motion of the constituent molecules of the group—a condition which holds in the case of the torsional vibrations of a fine metallic wire to the free end of which is attached a mass of great moment of inertia—it seems certain that the energy dissipated by true viscosity will be small in comparison with the energy dissipated in the breaking down of molecular groups (as in Maxwell's theory of a molecularly constituted solid), should such rupture take place to, possibly, a small extent only.

Wiedemann regards the loss of energy as due to the work done in rotating the molecules from one position of stable equilibrium to another, that rotation causing, in its turn, the permanent relative displacement of the molecules which occurs in the 'set': and this, of course, is quite a plausible hypothesis. In the homogeneous permanent torsional strain the final configuration of a given group of molecules is essentially similar to the initial configuration; yet we may, since a position of unstable equilibrium has been passed, say that the old configuration has broken down and that a new one has been formed; so that the loss of energy may be ascribed to the work done in breaking down molecular configurations. [Compare the explanation of the artificial twinning of a crystal of Iceland spar given in § 120.]

There does not appear to be any necessity to ascribe such a special type to the rupture of the configurations. It might rather be (at least in the case of large oscillations) that the permanent molecular rotation is a consequence of the permanent set which ensues on complete rupture of configurations—*i.e.*, rupture in which molecules in given configurations part company, new groups being formed. It appears to be most probable that the loss of energy in this total rupture greatly exceeds, group by group, the loss due to the cause which Wiedemann contemplates.

Such total rupture takes place in all liquids and in plastic solids; and, by analogy, we conclude that it takes place in all solids to a greater or less extent. Indeed, we do not need to rest upon analogy. The evaporation of ice below the freezing-point is due to such rupture; and so, probably, as has been remarked, is the characteristic smell of metals, such as freshly-cut copper or iron. Again, Spring has proved that metals, such as copper, sublime at temperatures far below their melting-points. At a given temperature, the molecules in a given configuration are in rapid motion with a certain average speed, the actual speed of a molecule being sometimes greatly in defect, sometimes greatly in excess, of this average. When the excess is sufficiently great, the molecule will pass beyond the range of the molecular action of the remaining constituents of the group, rupture of the configuration will occur, and the potential energy of deformation will be transformed into heat energy. In a sufficiently long interval of time such rupture *must* occur, according to the kinetic theory. In a given state of strain it occurs more readily the higher the temperature; at a given temperature it occurs more readily the greater the strain.

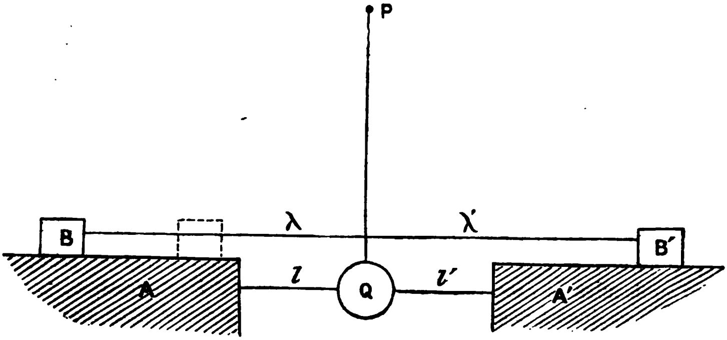


FIG. 52.

It is easy to see that, if a molecule is under directional control by molecular forces, rotation will in general ensue on rupture of a given group. A dynamical illustration will make this plain.

Let PQ represent a pendulum, with a massive bob Q, supported at the point P. Let the pendulum be situated symmetrically between the rigid fixed blocks A, A′, to which it is attached by equal and similar elastic cords l, l'; and let the pendulum be attached also by equal and similar elastic cords λ, λ′, to the rough movable blocks B, B′, which rest, symmetrically, with regard to P Q, upon A, A′. So long as B, B′ are not displaced, the equilibrium position of PQ will not be altered; but if the bob be moved to the right to such an extent as to move the block B into the position indicated by the dotted lines, so that the cord λ becomes slacker, the equilibrium position will change. The slackening of the cord λ corresponds to the removal of a molecule from a given configuration. The displacement of the equilibrium position corresponds to the set of the molecule PQ. There is change of relative position of Q, A, A′, and B′, and there is also rotation of PQ from its old direction. The change which takes place in one group is shared by all other groups which are in direct or indirect connection with that group. This could be shown in the model by attaching PQ by elastic cords to other pendulums. In this way the whole system takes a set.

Displacement of Q from the new position of equilibrium towards the right will take place strictly in accordance with Hooke's Law so long as the former maximum displacement is not exceeded. The old maximum force will be needed to cause the original maximum displacement to the right, but the actual displacement which it causes is now measured from the new equilibrium position. An equal force applied in the opposite direction shifts the equilibrium point back to its old situation, and gives a maximum displacement to the left, from this point, equal to the former maximum to the right. As regarded from the second equilibrium position to the right, the system resists more strongly displacements farther to the right than it resists equal displacements to the left. These facts illustrate Wiedemann's experimental results with regard to the torsion of wires.

The general result then, is that, in the case of small oscillations, the law of decrease agrees with that obtained on the assumption of true viscosity, while the dependence of the phenomenon upon the period of oscillation does not agree with the result obtained upon that assumption; so that the effect produced by true viscosity, should it exist to an appreciable extent, is overbalanced in one respect by an effect due to some other cause. Again, when the oscillations are large, the absorption of energy seems to be due mainly to change of molecular configurations—whether total change of the Maxwellian type, or rotational change of the type which Wiedemann assumes, or both. Most probably both occur; yet, when the oscillations are excessively large, it seems likely that the effect of the former will preponderate.

113. *Cohesion.—Tenacity.*—Cohesion is in general much more powerful in solids than in liquids. As an example of the cohesion of distinct solids, we have already had (§ 67) the case of two leaden

bullets with freshly-cleaned surfaces. Barton has worked copper surfaces so true that, when placed in contact, the lower mass may be lifted by means of the upper; and Whitworth has similarly worked steel surfaces. Another example occurs in the compression, by hydraulic means, of powdered graphite into solid blocks.

The parts of any body are kept together both by the force of cohesion and by mutual gravitation. In small bodies, such as stones, the part played by gravitation is totally negligable in comparison with that due to cohesion. But in large masses, such as the earth, gravitation has much the larger effect.

The force of cohesion has generally been regarded as a molecular attribute distinct from gravitation. But Lord Kelvin has pointed out that cohesion may be explained by means of the gravitational law.

If a bar of lead be cut into two parts, such that the freshly-cut surfaces accurately fit each other, the parts will readily reunite by cohesion when the surfaces are brought sufficiently close to each other. Such a phenomenon as this could not occur if matter were continuous and of uniform density throughout. For the range of the molecular forces (§ 121) is so small that only an extremely small amount of matter at one surface of the bar could sensibly attract, according to the gravitation law, a given particle at the opposed surface. But, in order to get comparatively a very large mass at one surface within 'molecular range' of the given particle at the other, we have only to suppose that, as regards density, matter is intensely heterogeneous on an invisibly small scale; and thus the molecular gravitational force might be sufficiently great to account for cohesion.

114. The property of *tenacity*, in virtue of which there is resistance to the drawing asunder of the parts of a body by tension applied in one direction, is obviously cohesion regarded from a different point of view. We may measure the tenacity of a substance by the tension which is required to rupture a rod or wire of that substance whose cross-sectional area is unity. The following table gives its values, for *sudden* rupture, in a few substances. Rupture will be *slowly* produced by somewhat smaller (frequently considerably smaller) tensions. The numbers in the table can be used for purposes of rough calculation only, as the property varies with the physical treatment and chemical purity of the substance. They represent the number of kilogrammes whose weight will produce sudden rupture of a rod which has a sectional area of one square millimetre.

Lead	2·3	2
Tin	4·2	3·6
Gold	28	11
Silver	30	16
Copper	41	32
Iron	65	50
Steel	99	54
Oak	10	
Ash	12	

The numbers in the first and second columns refer to the drawn and annealed states of the metals respectively. The tenacity of wood is of course very much smaller when the length of the rod is taken across the grain than when it is taken (as above) in the direction of the grain.

To reduce these numbers to their values in dynes per square centimetre, we must multiply by the factor 981 (10^5).

CHAPTER X.

THE CONSTITUTION OF MATTER.

115. EARLY in the history of science, discussion arose regarding the possibility of the infinite divisibility of matter. A drop of water may be subdivided into smaller drops of water; but this process cannot be indefinitely continued, for a point is ultimately arrived at beyond which subdivision cannot be carried without alteration of the chemical nature of the substance. But the problem with which we are now concerned goes deeper than this: we wish to know whether or not we would ultimately reach an indivisible part, or atom, of matter.

No answer can be given yet to this question, though various hypotheses have been framed regarding the ultimate constitution, or structure, of matter.

One of the most famous of these hypotheses is known as the *Lucretian Hypothesis* of hard atoms.

According to Lucretius, hard atoms exist; and they are indivisible because they are infinitely hard. The reason which he gave for their existence was that there must be a limit to the decay of matter—which he asserted to be a more rapid process than the agglomeration, or building up, of matter. If there were no such limit, all matter would, he said, have disappeared in the course of infinite past ages. His hypothesis may be true, but his assumption, upon which it was based, is not true; for we know that the building up of matter into larger parts is a more rapid process than its disintegration.

Lucretius further asserted that there must be void spaces between the atoms, otherwise motion of solids in fluids, such as that of a fish in water, could not occur. Here, again, his reason is not conclusive, although his conclusion is correct. Motion of solids in fluids could occur even if the hard atoms were closely packed together, provided only that they were smooth. This we know by the principle of fluid circulation in re-entrant paths. Some of the atoms would flow in behind the moving body as others were displaced in front. But the fact that all matter is compressible shows that there must be intervals between the hard atoms.

116. The atomic hypothesis of *Boscovich* is a mathematical device, which enables us to avoid some of the physical difficulties of the problem. Boscovich looked upon the atoms as mathematical points endowed with inertia and with attractive or repulsive forces varying with the distance. These forces are always attractive when the distance exceeds a certain superior limit, and are always repulsive when the distance falls short of a certain lower limit. They become infinite when the distance vanishes, so that no two atoms can occupy the same position at the same time. This theory may be worked out so as to determine the properties of a continuous medium so constituted. It has recently been developed by Lord Kelvin, who states that 'without accepting Boscovich's fundamental doctrine, that the ultimate atoms of matter are points endowed each with inertia and with mutual attractions or repulsions dependent on mutual distances, and that all the properties of matter are due to equilibrium of these forces, and to motions, or changes of motion, produced by them when they are not balanced; we can learn something towards an understanding of the real molecular structure of matter, and of some of its thermodynamic properties, by consideration of the static and kinetic problems which it suggests. Hooke's exhibition of the forms of crystals by piles of globes (§§ 119, 120), Navier's and Poisson's theory of the elasticity of solids, Maxwell's and Clausius' work in the kinetic theory of gases (Chap. XI.), and Tait's more recent work on the same subject—all developments of Boscovich's theory pure and simple—amply justify this statement.' 'It is also wonderful how much towards explaining the crystallography and elasticity of solids, and the thermo-elastic properties of solids, liquids, and gases, we find without assuming more than one transition from attraction to repulsion.' Poisson's theory of the elasticity of solids led to the conclusion that the compressional rigidity of a homogeneous isotropic (*i.e.*, with properties alike in all directions) solid should bear to its distortional rigidity the ratio 5/3. In Green's theory the ratio may have ny positive value from zero to infinity, or any negative value less than 4/3. Lord Kelvin shows that the simplest possible homogeneous assemblage of Boscovichian atoms realizes Poisson's medium, and that the second-simplest possible assemblage realizes Green's medium.

Some of the developments of the theory on the kinetic side will be given in the following chapter.

117. The most modern atomic hypothesis—and the one possessed of the greatest interest at the present day—is the vortex-atom hypothesis of Lord Kelvin, based upon the classical researches of Von Helmholtz on vortex motion. According to Kelvin, matter consists of the rotating parts of a perfect (§ 57), inert and incompressible fluid, which fills all space.

Since the fluid is perfect, any part of it, being once set in motion, will remain in motion for ever, and will be completely distinguished from all other parts. Thus the principle of conservation of matter is an essential part of the hypothesis. In this respect Lord Kelvin contrasts his theory with that of Lucretius in the following terms: 'The only pretext seeming to justify the monstrous assumption of infinitely strong and infinitely rigid pieces of matter . . . is that urged by Lucretius and adopted by Newton; that it seems necessary to account for the unalterable distinguishing qualities of different kinds of matter. But Helmholtz has proved an absolutely unalterable quality in the motion of any portion of a perfect liquid, in which the peculiar motion which he calls "wirbel-bewegung" (vortex-motion) has been once created. Thus, any portion of a perfect liquid which has "wirbel-bewegung" has one recommendation of Lucretius' atoms—infinitely perennial specific quality.'

The ultimate parts or vortices are indivisible, not from being perfectly hard, but because it is impossible to get at them in order to divide them.

The properties of such vortices may be illustrated by means of smoke-rings, such as those which are occasionally seen issuing from the funnel of a locomotive when the door of the furnace is suddenly closed, or those which are at times blown from the mouth of a cannon. These smoke-rings may readily be produced by means of a box, one end of which is flexible and the other end of which is perforated by a circular opening three or four inches in diameter. If a strong solution of ammonia be sprinkled on the floor of the box, and if a vessel containing a strong solution of hydrochloric acid gas (or, preferably, common salt on which strong sulphuric acid is poured) be also introduced into it, dense white fumes of ammonium-chloride are formed. A slight blow on the flexible end of the box is sufficient to drive out a portion of the air from the interior. This part revolves round and round, in the manner which is indicated in section in Fig. 53, and forms a complete circular vortex, the course of which may be traced for some distance through the surrounding still air. As the ring advances, its speed diminishes, and its diameter increases; and at last it disappears, its motion being stopped by friction.

If a second ring be projected after the former, with slightly greater speed and nearly in the same line, actual collision will not ensue, but each will move aside vibrating as an elastic ring would after direct impact. If it be projected exactly in the same straight line, the second one will grow smaller as it approaches the first, and, its speed increasing, will pass through the first. If the relative speed of approach be not too great, the two rings will not separate. Each will continue to revolve round the other, being alternately behind and in front of it. This corresponds to molecular or chemical combination of vortex atoms.

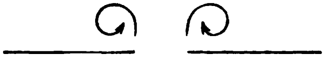


FIG. 53.

Currents are produced in the surrounding fluid which flow forwards in the direction of motion through the interior of the ring and pass back round the outside of it. It is the action of these currents which prevents actual contact between two impinging rings, and which makes it impossible to divide a ring formed in a perfect fluid.

The circular atom is the simplest which can exist, but a vortex-atom may be built up of any number of simple, or knotted, rings, linked or locked together in any manner whatsoever.

The mathematical difficulties which beset the investigation of the motion, and the mutual action, of vortices are so great that they have been overcome as yet only to a very slight extent. But, in so far as they have been overcome, the results are not unfavourable to the hypothesis. In this connection Maxwell remarks that 'one of the first, if not the very first, desideratum in a complete theory of matter is to explain first mass and then gravitation.' He then points out that, although inertia is postulated for the vortex medium, the mass of *bodies* requires explanation. 'It is true that a vortex ring at a given instant has a definite momentum and a definite energy, but to show that bodies built up of vortex rings would have such momentum and energy as we know them to have is in the present state of the theory a very difficult task.' Gravitation may be explainable by means of extra-atomic vortices. A modification of the theory, in which the fluid is not supposed to entirely fill space, has been advanced by Professor Hicks with the object of affording a more ready explanation of gravitation.

The great recommendation of the hypothesis is that it postulates nothing but the inert vortically moving fluid. All the known properties of matter are to be deduced from that one postulate, whereas other hypotheses assume the existence of special inter-atomic forces, and if one assumed set fails to produce required results, a new set can be conjured up to suit the case. As Kelvin said in 1867 : 'Probably the beautiful investigations of D. Bernouilli, Herapath, Joule, Krönig, Clausius, and Maxwell, on the various thermodynamic properties of gases, may have all the positive assumptions they have been obliged to make as to mutual forces between two atoms and kinetic energy acquired by individual atoms and molecules, satisfied by vortex rings, without requiring any other property in the matter whose motion composes them than inertia and incompressible occupation of space. A full mathematical investigation of the mutual action between two vortex rings of any given magnitudes and velocities, passing one another in any two lines, so directed that they never come nearer one another than a large multiple of the diameter of either, is a perfectly solvable mathematical problem; and the novelty of the circumstances contemplated presents difficulties of an exciting character. Its solution will become the foundation of the proposed new kinetic theory of gases. The possibility of founding a theory of elastic solids and liquids on the dynamics of more closely-packed vortex atoms may be reasonably anticipated. It may be remarked, in connection with this anticipation, that the mere title of Rankine's paper on "Molecular Vortices," communicated to the Royal Society of Edinburgh in 1849 and 1850, was a most suggestive step in physical theory' (see § 228).

Further considerations of the vortex theory will be made in the next chapter and in Chap. XXXII.

118. Standing in distinct contrast with these atomic hypotheses, we have the hypothesis that matter is continuous but intensely heterogeneous. The atomic hypotheses may be very roughly illustrated by a brick wall built without mortar. The separate bricks represent the various atoms, and there are gaps between them. The hypothesis which we now consider is similarly illustrated by a wall in which the gaps are filled up with cement. Viewed from a distance, the whole seems homogeneous (as matter does even when we inspect it with the most powerful microscopes), but if sufficiently closely inspected, it is seen to be heterogeneous.

Heterogeneity is a necessity whether atoms exist or not. This is indicated by many physical phenomena, notably, as Cauchy showed, by the dispersion of light, which, if the undulatory theory be true, could not occur were matter homogeneous and continuous—an idea advanced by Anaxagoras.

119. Although, as we have just seen, matter, *in its ultimate structure*, is essentially heterogeneous, yet it is customary and correct to speak of homogeneous matter. The whole thing depends upon the linear scale which we adopt in testing the homogeneity. A stone wall, while intensely heterogeneous with respect to a beetle, may be much too homogeneous with respect to a man, who wishes to climb it with ease. According to Kelvin and Tait, 'a body is called homogeneous when any two equal and similar parts of it, with corresponding lines parallel and turned towards the same parts, are indistinguishable from one another by any difference in quality.' Kelvin remarks that this is essentially the definition of crystalline structure if the 'part' of the body referred to does not contain an enormously great number of molecules. He then says that it seems impossible that equilibrium such as that which is manifest in a solid can be associated with an irregular random crowd of molecules; so that a homogeneous isotropic solid is but 'a solid composed of crystalline portions having their crystalline axes or lines of symmetry distributed with random equality in all directions. The proved highly perfect optical isotropy of the glass of object-glasses of great refracting telescopes, and of good glass prisms, seems to demonstrate that the ultimate molecular structure is fine-grained enough to let there be homogeneous crystalline portions, which contain very large numbers of molecules, while their extent throughout space is very small in comparison with the wave-length of light.' If, therefore, we can imagine a medium, composed of molecules exerting mutual forces, which simulates in its essential structure the known forms of crystalline substances; and if, further, this medium possesses the elastic qualities of actual solid media, we shall have strong presumptive evidence in favour of the molecular theory of matter. This evidence will become practically conclusive if we find that, by motion of these molecules, we can explain the kinetic phenomena of matter in its solid, liquid, and gaseous forms (see Chap. XI.).

Now it is a fact that if we assume the existence of molecules which exert action in all directions throughout a small range, we can explain the various phenomena of crystalline structure. Let us suppose first that the molecules are in stable equilibrium under their mutual action when nearest neighbours are at a definite distance apart, and that they exert equal forces in all directions. We can obviously make a model of such an arrangement, as did Hooke, by means of equal spherical balls or marbles, a molecule being sup-

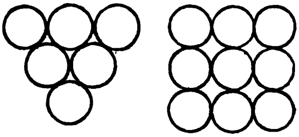


FIG. 54.

posed to be situated at the centre of each, and the whole system forming a closest-packed homogeneous assemblage.

We may lay down a plane triangular arrangement of marbles (Fig. 54), and then build up a solid by placing a second set of marbles above the first, so that each marble in the second set fits into the hollow between three in the first set, and so on—the number placed in each layer being less than the number in the layer beneath by the number contained in the edge of that lower layer. In this way we form a regular tetrahedron.

This tetrahedron has six edges, and if these edges be pared off by planes equally inclined to the adjacent faces, we ultimately get a cube.

The cube has eight vertices, and if these vertices be pared off by planes equally inclined to the faces which meet at these vertices, a regular octahedron is obtained.

If the twelve edges of the cube be bevelled by planes equally inclined to the adjacent faces, a rhombic dodecahedron is ultimately produced.

It is by no means difficult to account in this way for all the various symmetrical forms of crystals. belonging to the regular system; and if we replace the component spheres by ellipsoids o revolution, we can account, by similar processes, for all known crystalline forms.

From such facts, even by themselves, we are led to believe that crystalline bodies are built up of particles which set themselves, under the action of their mutual forces, in the position of least potential energy, *i.e.*, in the position of stable equilibrium.

120. The square order of arrangement (indicated in the above figure) in which any one sphere touches four spheres in the layer immediately below, is in no way different from the triangular arrangement just considered; for, if we remove an edge row o

spheres from the triangular pyramid, it is at once evident that the particles are arranged in square order in planes which bevel an edge symmetrically. The triangular arrangement is exhibited in the faces of the square pyramid. The identity of the two arrangements may be still further seen if we build up the triangular pyramid by a different process. In each layer, symmetrically arranged with regard to the layer beneath, let the number of marbles be equal to the number which would be contained in the layer beneath if the marbles contained in *two* of its edge rows were removed. The three faces of the pyramid will be found to be at right angles to each other, so that the pyramid represents a corner of a cube; and the square arrangement of marbles will be exhibited in the faces.

In both the triangular and the square order of arrangement any one particle touches twelve others. In the triangular order, a particle touches six others in its own layer, and three in each of the adjacent layers. In the square order, a particle touches four in each of these layers.

It is true indeed that all the crystalline forms of the cubic system may be explained by other methods of arrangement, but these are of little interest to us physically, as they do not correspond to that position of most stable equilibrium which free, mutually attractive, particles would naturally assume. Thus we might build up a cube first of all on the square order of arrangement—a particle in any one layer resting on the top of one in the subjacent layer; and from this cubic form we might then, as above, produce the other crystalline forms. Again, we might start with an open square arrangement (indicated in Fig. 55), the particles in any one layer being so far apart that another particle which is set in the hollow between four will just touch a particle which is set in the corresponding hollow on the other side of these four. In this way also all the various crystalline forms of the regular system may be produced. In the former method, any particle touches four particles in its own layer, and one particle in each of the adjacent layers—six in all; in the latter method, a particle touches none in its own layer, four in

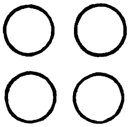


FIG. 55.

each of the layers immediately preceding and succeeding its own, and one in each of the layers outside of these—ten altogether. In neither of these cases is the equilibrium so stable as in the case which was considered in last section.

A different closest-packed arrangement has been described by W. Barlow. In the triangular arrangement described in § 119 the marbles in the third layer are not over those in the first. In the arrangement now considered they are so placed. Similarly those in the fourth layer are over those in the second, and so on. While, in the planes of these layers, the lines of contact of marbles are rectilinear, spiral lines of contact run downwards through the structure. Barlow points out that this arrangement may be exemplified in quartz, in which a spiral structure is known to exist from its power of rotating the plane of polarisation of light. As Kelvin remarks, the arrangement is not homogeneous, but consists of two homogeneous groupings, which are exhibited in the set of even layers and the set of odd layers respectively.

We have now to discuss the question of the elastic qualities of such a medium. Consider the cube with the square arrangement exhibited in the faces (see top of p. 129). Lord Kelvin has shown that, with an arbitrary law of molecular force, the rigidity relative to shear parallel to the faces of the cube is not equal to the rigidity relative to shear parallel to the diagonal planes of the cube. By proper choice of the law of force it is possible to make the two rigidities equal, and Kelvin finds that then the distortional rigidity must be three-fifths of the compressional rigidity—a condition which resulted also from Poisson's theory of elastic solids, but which does not in general hold true in nature. Hence it follows that this closest-packed homogeneous assemblage of Boscovichian atoms must be put aside as inadequate. But Kelvin has shown that a similar double atom arrangement may be found which is free from the objection. Consider the above assemblage of single atoms. Draw through the middle point of each line joining nearest neighbours a plane perpendicular to it. These planes divide space into rhombic dodecahedra, having single atoms at their centres. Choose four of the eight points at the trihedral angles of a dodecahedron in such a way that the lines joining the centre to each pair of the four points make equal obtuse angles with each other. Select any one of these points and join it to the centre. Then draw lines, equal and parallel to this joining line and in the same direction, from the remaining three points. These lines will terminate at the centres of neighbouring dodecahedra. Hence new single atoms placed at the four given points of each dodecahedron will form with the old atoms a closest-packed homogeneous assemblage of double atoms. This assemblage (and each similar multiple-atom assemblage) is free from the objection: it constitutes the second-simplest assemblage referred to in § 116 as realizing Green's medium.

It only remains to consider *how* it is that a crystal grows in its definite geometrical form with plane surfaces. The potential energy of a molecule in the interior is an absolute minimum when it has got as close as it can to the greatest number of the adjacent molecules. But this condition does not hold at the surface. The potential energy of a molecule at the surface reaches the least possible value when the surface molecules are as closely grouped as possible; and, as the arrangement of the molecules is uniform throughout any given plane, the surfaces must be those plane surfaces in which the grouping is closest. And planes parallel to these will be planes of cleavage; for the substance will tend to divide in such a way that the new surfaces exposed will have the least energy. At an edge or corner the resultant of the molecular forces acts inwards, and so edges and corners tend to be produced symmetrically, in order that equilibrium may ensue. In the regular system the standard form will be octahedral, for the surfaces of the regular octahedron are surfaces of closest grouping of homogeneously closely-packed spheres. Edges may be bevelled, or corners truncated, if the increase of energy, caused by wider grouping in the new surface so formed, is balanced by the decrease of energy due to the diminution of total surface. If spheroids be used instead of spheres, the standard form may be cubic, dodecahedric, or pyramidal, etc. Again, the surface energy depends upon the medium in which the crystallisation takes place, which explains the known change of crystalline form which sometimes occurs when the medium is changed (compare § 102). These considerations were first stated by Liveing.

The ready explanation which the above considered principles give of a remarkable property of Iceland spar is well worthy of note. The form of a crystal of this substance is represented in § 208. Baumhauer has shown that, if pressure acting towards an obtuse-angled corner be applied symmetrically, by means of a knife, at a point on one of the edges meeting at that corner, the upper portion of that part of the crystal in front of the knife splits away from the part behind and shears forwards, so as to form with the under portion of the front part a complete twin crystal. Now, Huyghens, in 1690, and (independently) Wollaston, in 1812, had suggested the following structure for Iceland spar. If the equilateral tetrahedron, described in § 119, suffers uniform shrinkage perpendicular to its base in the ratio of $\sqrt{8}$ to 1, its acute trihedral angle will become the obtuse trihedral angle of Iceland spar. The spheres will become ellipsoids of revolution, with their short diameter parallel to the optic axis of the crystal; and we have thus a system whose mechanical and optical properties simulate those of the spar. Refer now to Fig. 25, § 40. If *a*, *b*, *c*, *d* represents the crystal, the ellipsoids lie with their short diameters in the direction *db*. If they can be turned round, sliding over each other till they take up a new position of closest packing, *a'*, *b'*, *c*, *d* will represent the resultant form with the short axes of the ellipsoids directed parallel to *ca'*. Artificial twinning would thus be produced.

Kelvin also independently thought of this explanation, but finally rejected it, because of the great amount of internal commotion which would be necessitated in the process of shearing, in favour of the following. Start with the second form of the triangular arrangement described at the commencement of the present section. Shrinkage to only half the extent of that needed in the previous case is now required to produce the form of the spar, and much less internal commotion is necessitated in the process of shearing.

121. Many other physical phenomena, besides those which are exhibited by crystalline bodies, make it a practical certainty that matter is molecular in its structure, and also enable us to obtain approximate estimates of the range of the molecular forces, and of the average distance between contiguous molecules.

One such class of phenomena is that exhibited by liquid films. We have seen (§ 101) that the existence of molecular forces would account for surface-tension. And we have seen also (§ 102) that the surface-tension remains practically constant until the thickness of the film is very largely reduced. There are optical methods (§§ 192-195) by which the thickness can be very accurately ascertained. Reinold and Rücker have shown by these methods that the surface-tension of a soap-bubble begins to diminish when the thickness is between 96 and 45 micro-millimetres (millionths of a millimetre, one inch being about equal to 25·4 millimetres). It diminishes to a minimum when the thickness is 12 micro-millimetres, and then increases again to a maximum.

Plateau had previously shown that the tension is unaltered when the thickness is reduced to 118 micro-millimetres, and he concluded that the range of molecular forces is less than 59 micro-millimetres. Maxwell, however, has given theoretical reasons for the belief that the tension may not alter until the total thickness of the film is equal to the range of the forces, which would make Plateau's superior limit 118 micro-millimetres.

By measurements of the height to which a liquid rose, because of the so-called capillary forces, between parallel glass plates which were coated with very thin wedge-shaped metallic films, Quincke was led to the conclusion that the forces between the glass and the liquid became evident when the thickness of the metallic film was 50 micro-millimetres.

Wiener has found that the phase of the vibrations of light (§ 217) which is reflected from a thin film of silver deposited on mica begins to alter when the thickness is 12 micro-millimetres, and that it is possible to detect the presence of a film of silver not exceeding 0·2 micro-millimetre in thickness.

Estimates of the magnitude of the molecular range have also been based upon the thickness of the films of gas which are condensed upon solids, but these are open to great objection.

From all these results we may conclude that the order of magnitude of the range of molecular forces is about 50 micro-millimetres, or one five-hundred-thousandth part of an inch.

122. It is possible, also, to obtain an estimate of the coarsegrainedness of matter; that is, of the average distance between molecules.

Two plates of different metals, say copper and zinc, if placed in contact, become each electrified oppositely, and so exhibit mutual attraction (see Chap. XXIV.). If the plates have each an area of one square centimetre, and be at a distance of one hundred-thousandth of a centimetre apart, they will, when joined by a metallic connection, attract each other with a force of two grammes weight. Hence the work done in bringing them into this position by means of the electric attraction alone would be 2/100,000ths of a centimetre-gramme. If we now build up a cube of such pieces of metal, alternately zinc and copper, the thickness of each being 1/100,000th of a centimetre, and the distance apart of each pair being 1/100,000th of a centimetre, the work done by the electric attraction is 2 centimetre-grammes. If this work were spent in heating the mass of metal, the temperature would rise by less than the sixteen-thousandth part of a degree centigrade. But if the thickness of the plates and their distance apart were 1/100,000,000th of a centimetre, the work would be sufficient to raise the temperature of the mass by nearly 62° C. And, if the plates and the spaces between them were made yet four times thinner, the work would produce more heat than is given out by the molecular combination of zinc and copper. Hence the magnitude of the contact-electrification of zinc and copper must sensibly diminish before the substances are so finely divided as we have supposed. But this suggests that there the surface tension until there are considerably fewer molecules in the edge. On the other hand, if we assume that the true value is given by Kelvin's lower limit, we find that there are only six molecules in the edge—a condition which seems to be as improbable as the former. Again, when the subdivision has been carried to this extent, the numerical value of n has already become nearly one million—a number which is sufficiently near the experimental values above given to add considerable weight to the general reasoning. In fact, the weight of the combined evidence of all the methods is very great.

CHAPTER XI.

THE KINETIC THEORY OF MATTER.

123. THE first glimmerings of the idea that the observed properties of matter may be due to motion occurred as far back as the time of Democritus and Lucretius. But the idea did not develop into an actual physical hypothesis until Hooke, and, later, Daniell Bernouilli suggested that gaseous pressure may be due to the impact of the molecules of the gas upon the sides of the vessel which contains it. Somewhat later, Le Sage, as we have already seen, applied the same principle to the explanation of gravitation, and various developments were made by Prevost, Herapath, and Waterston. In 1848, Joule calculated the speed which the particles of a given gas must have in order to produce a given pressure. But the full mathematical development of the Kinetic Theory of Gases is due mainly to Clausius and Maxwell.

124. In the kinetic theory it is supposed that the particles of a gas are darting about in all directions with great average speed. Some of the particles may, for a short time, have very much smaller speed than this average—may indeed be at rest for a moment; and others may be moving with much greater speed than the average. This average is the square root of the mean of the squares of the individual speeds of the various molecules, and is called the *mean square speed.*

Collisions are supposed to occur amongst these particles. In the interval between any two successive collisions of a particle there is a certain average distance which the particle describes, and this distance is called the *mean free path.* The mean free path, under ordinary conditions, is large in comparison with the diameters of the molecules, which are regarded as being smooth hard spheres with unit co-efficient of restitution.

The collisional force between two molecules is assumed to be repulsive, just as it would be in the case of elastic solids; and, as a first approximation, the molecules are not supposed to exert any mutual force except at the instant of collision. But, actually, in many, or even most, of the so-called 'collisions,' true contact may not occur. The real forces may be attractive, and two rapidly-moving molecules coming within the range of mutual attraction may whirl round each other in sharply-curved paths, as a comet dashes round the sun, the result being the same as if actual contact had taken place. Some contacts must occur unless the molecules are infinitely small, which we cannot admit.

Experiments made by Joule and Kelvin on certain gases show that, if the density of each be varied while its total energy remains constant, the temperature is somewhat higher when the density is greater. Hence, if (§ 126) equality of temperature in two portions of a gas means equality of average kinetic energy per molecule, it follows that the potential energy is somewhat less in the denser condition. But this indicates molecular attraction at the average distance of the molecules experimented with.

125. *Gaseous Pressure.—Boyle's Law.*—Let n be the number of molecules per unit volume which are moving in a given direction with speed which differs extremely little from a certain quantity u. The number of such molecules which pass per unit time across unit area drawn perpendicular to the direction of motion is nu; and, if m be the mass of each molecule, the momentum which these particles carry with them is mnu^2. By Newton's Second Law of Motion, this must be equal to the pressure produced by such molecules on the side of the vessel. Hence, *the square of the speed being involved*, we see that, so far as the pressure is concerned, we may assume that each particle is moving with the mean square speed.

The total pressure per unit of area in the direction considered is therefore $m\mathrm{N}\overline{u^2}$, where N is the total number of particles per unit volume, and $\overline{u^2}$ is the mean value of u^2 for all the molecules. Similarly, the pressures per unit area in the two remaining mutually rectangular directions may be written $m\mathrm{N}\overline{v^2}$ and $m\mathrm{N}\overline{w^2}$ respectively. But, in a gas which is at rest as a whole, all these quantities are equal; and so we have for the pressure p the expression $p=\frac{1}{3}m\mathrm{N}(\overline{u^2}+\overline{v^2}+\overline{w^2})$, which we may put in the form

$$p=\tfrac{1}{3}m\mathrm{N}\overline{\mathrm{V}}^2=\tfrac{1}{3}\rho\overline{\mathrm{V}}^2 \quad . \; . \; . \; . \; . \; . \quad (1)$$

where $\overline{\mathrm{V}}$ is the mean square speed independent of direction and ρ is the density of the gas. By means of this result, Joule calculated the value of $\overline{\mathrm{V}}$ in various gases. In hydrogen it is somewhat over 1,800 metres per second. Values for three other gases are given in § 129.

If (§ 126) $\overline{\mathrm{V}}$ is constant when the temperature is steady, this equation asserts that the density at constant temperature is directly proportional to the pressure—which is Boyle's Law.

The vortex-atom theory also leads to these results. On the question of gaseous pressure, Lord Kelvin says: 'When a vortex-ring is approaching a plane, large in comparison with the dimensions of the ring, the total pressure over the surface is nil. When a ring approaches such a surface, it begins to expand, so that if we consider a finite portion of the surface, the total pressure upon it due to the ring will have a finite value when the ring is close enough. In a closed cylinder, any vortex-ring approaching the plane end will expand out along the surface, losing in speed as it so does, until it reaches the cylindrical boundary, along which it will crawl back on rebounding to the other end of the cylinder. As it approaches, it will therefore exert upon the plane surface a definite outward pressure whose time-integral is equal to the original momentum of the vortex, and a precisely equal pressure as it leaves the surface. Hence, in the case of myriads of vortex-rings bombarding such a plane surface, though no individual vortex-ring leaves the surface immediately after collision, for every vortex-ring that gets entangled in the condensed layer of drawn-out vortex-rings another will get free, so that, in the statistics of vortex-impacts, the pressure exerted by a gas composed of vortex-atoms is exactly the same as is given by the ordinary kinetic theory, which regards the atoms as hard elastic particles.'

Professor J. J. Thomson finds that the vortex theory leads to Boyle's Law as an approximation, and that it 'explains the deviation of gases from Boyle's Law. In this respect it compares favourably with the ordinary theories.'

126. *Avogadro's and Charles' Laws.*—The equation (1) may be written

$$pv=\tfrac{1}{3}\overline{V}^2 \; . \; . \; . \; . \; . \; . \; (2),$$

where v is the reciprocal of ρ, *i.e.*, it is the volume of unit quantity of the gas. If we compare this with the expression (§ 240) $pv=RT$, we see that the mean square of the speed of molecular motion is proportional to the absolute temperature as measured by a gas thermometer (§§ 240, 241) filled with the particular gas under consideration.

It follows from the principles of the kinetic theory that, in a *mixture* of two gases in equilibrium, the average kinetic energy per molecule of each gas is the same. And if we assume the truth of Avogadro's Law that there is the same number of molecules in unit volume of each of two separate gases at a given temperature and pressure, (1) shows that in two such gases the average kinetic energy per molecule is identical. For it gives $\rho_1\overline{V}_1^2=\rho_2\overline{V}_2^2$, where the suffixes refer to the different gases, and this condition asserts that the average kinetic energy per unit volume has the same value in both gases; so that, dividing both sides of the equation by the (common) number of molecules per unit volume, we get the required result. Conversely, if we assume—as is done in the kinetic theory — that two gases at the same temperature have the same average kinetic energy per molecule, we can deduce Avogadro's Law as a result. But it must be carefully observed that the truth of Avogadro's Law does not establish the truth of this assumption: it merely proves that, on this theory, two gases at the same temperature *and pressure* have equal average kinetic energy per molecule. If, however, the gases rigidly obey both this law and Boyle's Law, the truth of the assumption follows: for then, in any one of the gases, the average kinetic energy per molecule remains constant, no matter how much the pressure may vary. To see this we only require to note that this average is the ratio of the pressure to the number of molecules in unit volume—which ratio, by Boyle's Law, is constant.

Equation (2) shows that any gas which obeys these laws will, at constant pressure, expand equally for equal increments of temperature, *provided that the temperature be measured by the average kinetic energy per molecule*, which is a generalisation of the above assumption. And, further, with this proviso, the equation also shows that any two such gases will expand proportionately as their common temperature rises. These two results constitute Charles' Law (§239).

The deviations from Boyle's and Charles' Laws can be explained on the kinetic theory if we take into account molecular action between the particles at greater than collisional distances.

127. *Diffusion.—Thermal Conductivity.—Viscosity.*—The question of gaseous diffusion has been already discussed in Chap. VII. The kinetic theory asserts that each particle is darting about with great speed, and is only prevented from moving rapidly away from the neighbourhood of its position at any given instant by means of the great number of collisions to which it is subjected by other particles. These collisions change the direction of motion of the molecule an immense number of times per second, and thus the diffusion of particles throughout a mass of gas is a very slow process. The law of inter-diffusion given by the theory is identical with that deduced from observation.

The diffusing molecules carry their kinetic energy with them, and there is interchange of energy during collision. This transference of energy constitutes (Chap. XXII.) the process of thermal conduction. And this process is slightly more rapid than the transference of the molecules themselves; for, though a molecule may be turned back by a collision, its energy is handed on by means of the molecule which collides with it.

If two portions of a gas are moving relatively to each other, the molecules of each inter-diffuse, and consequently there is interchange of momentum. But, on the average, the momenta of the particles of each portion are in opposite directions, and so the relative motion is gradually stopped. This explains gaseous viscosity. An illustration may be taken from two railway trains running on parallel lines past each other. If luggage were thrown constantly from each train into the other, the interchange of momentum resulting from the impacts might soon reduce the trains to relative rest.

128. *Evaporation, Dissociation, etc.*—As a gas becomes more and more condensed the mean free path of its molecules becomes smaller and smaller, until, in the liquid state, its magnitude is excessively small in comparison with the magnitude of the mean free path of a gaseous particle. Still, in the liquid, the molecular action is precisely of the same character as that in a gas. But, as a result of the comparative closeness of the molecules, the transference of energy by impact (that is, the conduction of heat) is a much more rapid process than the transference of the molecules themselves; and the rate of diffusion of the molecules of a liquid is very small in comparison with the rate of diffusion of gaseous molecules.

In a liquid some of the quickly-moving particles may escape from the attraction of neighbouring molecules and become particles of vapour. At the same time some vapour particles may become entangled in the liquid. When these two processes take place at the same rate, the liquid and its vapour are said to be in equilibrium. When the former process occurs most rapidly, the liquid is said to be evaporating; and when the latter process preponderates, the vapour is said to condense.

Dissociation occurs when an impact is so violent as to break up a compound molecule into its constituent parts. Probably dissociation occurs in all fluids to a slight extent, even when their temperature is far below the ordinary temperature of dissociation, but, in this case, recombination occurs as rapidly. As the temperature rises, the impacts become more violent and dissociation goes on at a greater rate, but not at so great a rate that recombination cannot occur as quickly. When, finally, the so-called temperature of dissociation is reached, recombination is unable to balance dissociation, and the substance is resolved into its components. If the temperature is again allowed to fall, recombination may, or may not, occur, depending on the condition whether or not energy will be degraded (§ 11) in the process. If more energy will be degraded by the occurrence of recombination than by its non-occurrence, recombination will ensue. [See § 254.]

An important result of the kinetic theory is that, in a vessel filled with a mixture of different gases, the final distribution of each gas under the action of gravity is the same as if no other gas had been present.

Another very important result (which, together with the former, is due to Clerk-Maxwell) is that a vertical column of gas, when in equilibrium under the action of gravity, has the same temperature throughout, or, in other words, gravity has no effect upon the conditions of thermal equilibrium. We may also obtain, very simply, the value of the greatest rate of diminution of temperature with height in the atmosphere which is consistent with stability. Let m be the mass of a molecule. If this molecule falls downwards through a height h, the increase of energy is mgh (§ 43). On the kinetic theory this must be proportional to the increase of temperature, t. Hence we get $mgh=kt$, where k is a constant. The quantity on the left-hand side of the equation is the gain of energy per molecule. Per unit of mass, the gain is gh. This enables us to find the value of k. For we have $pv=RT$, where T is the absolute temperature; and (§ 86) pv is energy per unit mass. Hence R is the factor which changes the measure of energy in terms of temperature into dynamical units. Therefore $k=mR$; and so $t/h=g/R$ is the equilibrium rate of variation of temperature. If the temperature fell off at a greater rate than this, the upper strata would become too dense, and so would descend.

Tolver Preston and Johnstone Stoney have suggested that the reason why the moon has no atmosphere may be that its gravitational attraction was insufficient to prevent the speed occasionally acquired by molecules from carrying them beyond the range of attraction. G. H. Bryan has calculated from the theory that, in the case of oxygen at 0° C, one molecule per three billion is moving fast enough to get away from the moon; while only one in every $2{\cdot}3\times10^{329}$ would escape from the earth's attraction.

129. An expression has been deduced from this theory by which we can calculate the length of the mean free path of a molecule in an ordinary gas, such as air, in terms of observed values of the viscosity, and of the material and thermal diffusivities. According

to Maxwell the value of this quantity in the case of hydrogen is about $9{\cdot}7(10)^{-6}$ centimetre. The theory also shows that the number of particles per cubic centimetre of ordinary gases is about $2(10)^{19}$, and that the diameter of a (supposed hard and spherical) molecule is about $6(10)^{-8}$ centimetre.

The following results are taken from Maxwell's data:

	Hydrogen.	*Oxygen.*	*Carb. Oxide.*	*Carb. Acid.*
Mean square speed at 0° *C. expressed in metres per second*...	1888	437	505	403
Mean free path (*unit* = 10^{-8} *cm.*)	980	569	490	384
Number of collisions per ten-millionth of a second ...	1775	765	949	972
Diameter (*unit* = 10^{-10} *cm.*) ...	580	760	890	940
Mass (*unit* = 10^{-25} *gramme*) ...	46	736	644	1012

The length of the free path increases as the density of the gas diminishes. Tait and Dewar first found that, in a good vacuum, it may amount to several inches. The action of Crooke's *radiometer* depends upon the great length of the free path. This instrument consists of four vanes of mica mounted at the ends of two light rods, which are fastened by their centres to a vertical axis which is free to rotate. Each disc of mica lies in the vertical plane which passes through the rod to which it is attached, and the two rods are placed at right angles to each other. Also each disc is blackened on one side and is bright on the other, the blackened sides being all similarly situated with regard to rotation around the vertical axis. The whole is mounted inside a glass vessel from which the air is largely extracted. Radiant heat (or light) falling upon the vanes is absorbed more freely by the black surfaces than by the bright surfaces, and so the former become hotter than the latter. Hence the particles of air which impinge upon them are driven off again more violently than are the particles which impinge upon the bright sides, and so, by the third law of motion, the necessary reaction results in rotation.

The exhaustion of the air is carried to such an extent that a particle which is repelled from a disc rarely encounters another particle before it strikes the side of the vessel. If the exhaustion is carried too far the effect is diminished, because the number of particles which strike a disc in a given time is lessened.

Gaseous matter whose molecules have very large free paths is sometimes called 'radiant' matter.

130. The statement that the diameter of one of these hypothetical (hard, smooth, spherical) molecules is 7 hundred-millionths of a

centimetre is not intended to imply that this is the size of an actual material molecule. It merely asserts that, on the average, this is about the least distance to which the centres of two molecules can approach during collision.

But, even if they be smooth and spherical, the molecules of matter cannot be hard; for, in this case, the time of describing the mean free path must vary inversely as the average speed of the particles. But, on the contrary, Maxwell's experiments on gaseous viscosity show that the time is independent of the speed. This would be possible with soft elastic particles.

Again, we have found that the unmodified theory leads to Boyle's Law as a necessity, while this law is not strictly observed by any real gas. Molecular attraction must be assumed in order to explain the observed deviations. Without assuming any special law of attraction, attempts have been made to find a more accurate expression for the relations amongst pressure, volume, and temperature in a gas, by the mere consideration that such attraction exists. The unmodified law indicates that the volume diminishes indefinitely under increasing pressure. D. Bernouilli first remarked that finite size of the molecules prevents this. Hence, instead of the volume in the expression for Boyle's Law, we must rather take the volume less a certain small quantity. Hirn introduced this correction, and also remarked that molecular attraction increases the total pressure, so that the actual pressure is the external pressure plus a certain quantity. He assumed that these two quantities were constant. Van der Waals gave theoretical reasons for the belief that the latter quantity should vary in inverse proportion to the square of the volume, and so deduced an expression which is in wonderful accordance with experimental results in many respects. Thus it leads to the result, shown experimentally for certain gases by Cailletet and Amagat, that the product of pressure and volume has a minimum value at a definite pressure: and it has a general applicability also to the phenomena of liquefaction. Tait, not agreeing with Van der Waals' reasoning, deduces a different expression, which he has tested by reference to experimental results and contrasted with Van der Waals' expression.

The following table gives, in the third column, values of the product of pressure and volume for air calculated by Van der Waals from his formula. The second column contains the true values as observed by Amagat, and the corresponding pressures are given in the first column, the unit being the weight of a column of mercury one metre in height:

1	1000	1000	64·00	—	977
9·94	994	—	68·55	979	—
19·77	998	—	72·16	—	978
24·07	—	984	78·67	984	—
29·53	984	—	84·22	—	980
34·90	—	982	89.03	989	—
39·25	981	—	99·85	998	—
45·24	—	978	101·47	—	990
48·86	977	—	111·09	1009	—
55·50	—	977	133·89	—	1009
58·64	977	—			

There are other, even more cogent, reasons why the molecules cannot be smooth, hard, and spherical. Thus the great complexity of the spectra (Chap. XV.) of many gases and vapours shows that the molecule is a very complex structure with a great many degrees of vibrational freedom. But a hard, smooth, spherical molecule has no more than three degrees of freedom—all translational. And again, the theory asserts that each additional degree of freedom which the molecule possesses requires that the ratio of the specific heat (§ 245) of the gas at constant pressure to that at constant volume shall be made larger; and the actually observed values of this ratio are far smaller than that indicated by the theory. Further, it seems certain that, in a gas whose molecules are small, elastic, solid bodies, the energy of translation of the molecules must gradually be converted into energy of vibration of smaller and smaller period, so that finally the particles would come to rest.

131. But these difficulties with which the theory is beset should not lead us to discard it. Its results, briefly indicated in part in the preceding sections, place it upon far too firm a basis; and we should rather inquire more closely into the truth or probability of the fundamental assumptions of the theory, and into the correctness of the deductions made from them. The difficulty regarding the specific heats seems to be due largely to a too rapid theoretical generalisation. And the difficulty regarding the transformation of the translational energy of elastic solids into vibrational energy does not, Lord Kelvin remarks, seem to apply to the case of fluid vortices.

132. We are not, however, to rest content with a kinetic theory of gases only. It is desirable, if possible, to recognise all the properties of matter as the results of motion of the parts of a medium in which nothing but incompressibility and inertia is postulated.

We can make, from rigid portions of matter, a complex which possesses elasticity. A gyrostat (essentially a heavy fly-wheel, of great moment of inertia, free to rotate about its axis) mounted on gymbals, but not set in rotation, serves very well as a model of a plastic body. It may be turned about into any position which we desire it to assume, and has no tendency to re-assume its original position. But, if the fly-wheel be set in rotation about its axis, the system at once acts as if it possessed rigidity and elasticity. If it is sharply struck, so as to suddenly turn its axis to a slight extent from its original direction, it will oscillate rapidly about, and finally come to rest in, its first position.

Again, by means of rigid rods on which revolving fly-wheels are pivoted, we can construct, as Kelvin showed, a framework, which acts like an elastic spring, and may represent an elastic molecule. By joining together millions of such arrangements, it is possible to construct a model of an elastic solid through which distortional waves will pass, and which, under proper conditions, will exhibit, with regard to these waves, the same phenomena as those which are shown when light passes through a medium placed in a field of great magnetic force.

Kelvin remarks that the vortex theory shows that all that can be done by such a model might be done by a model in which vortices in a perfect fluid take the place of the revolving fly-wheels. And further, with such vortex molecules we would have—what cannot be had with the rigid fly-wheel molecules—a model of a gas, whose molecules exert mutual force.

Such considerations support the idea that it may be possible to explain apparently statical properties of matter in terms of motion.

In conclusion, we may consider Lord Kelvin's applications of the Boscovichian form of the kinetic theory to the explanation of the various changes of state which matter undergoes.

If the positions and velocities of single Boscovichian atoms are distributed with average equality throughout space, the results of this form of the theory agree with those of the unmodified hard-sphere form, provided that none of the collisions take place amongst more than two atoms. If three atoms collide simultaneously, two may remain in combination under the action of their mutual forces, and possibly even all three may so remain. Yet, if the initial average speed is sufficiently great, there will be few such doublets or triplets; and those which do occur are liable to dissociation into single atoms by fresh collisions. This exemplifies a perfect monatomic gas at temperatures far above its critical point. On the other hand, if the initial average speed be sufficiently small, doublets will readily be formed as the result of triple collisions. The general tendency will be to send off the single atom with large speed. Hence, the centre of inertia of the newly-formed doublet may have considerable speed. The components of the doublet will also have in general great energy of relative motion because of the exhaustion of potential energy due to their mutual attraction. Future collisions with the (averagely) slow-moving atoms will diminish the speed of the double atom, and will also diminish its internal energy; so that it finally moves slowly with small internal motion. Collisions of two such doublets will in general give rise to a triplet moving with considerable speed and internal energy and a single atom moving with considerable speed; and subsequent collisions with slow-moving single atoms will tend to reduce the energies of absolute and internal relative motions, so that finally the triplet moves with slow speed and with its constituent atoms nearly in equilibrium at the corners of an equilateral triangle. Similar results hold in the case of collisions of multiple atoms of higher order. And, in the case of collision of a multiple atom with a single slowly-moving atom, the single atom will frequently remain in combination, the probability of its doing so being greater the greater the multiplicity of the other atom is.

The internal energy of the group is increased by the exhaustion of potential energy, as each single atom falls into it. But an atom will frequently be driven out of combination when a fresh one falls in. And the atom so driven out will in general go off with much greater speed than the atom which fell in possessed: for the atom which is most readily driven out is that one which is most nearly able to go off without external aid. In this way the average kinetic energy of vibration per atom of the group may be constant, while the exhaustion of potential energy is compensated by increased kinetic energy of the atoms in the atmosphere surrounding the group. If proper appliances be used, this excess of kinetic energy may be tapped off from the system, and we thus have a crystalline substance growing by condensation of a surrounding atmosphere from which latent heat is being abstracted.

If the abstraction of kinetic energy be stopped, the kinetic energy of the free atoms will gradually increase until a stage is reached at which as much energy is carried into the crystalline group by the falling in of a fresh atom as is carried out by the ejection of an atom from the group, and, on the whole, as many atoms will fall in during a given time as are ejected in that time. The entire system is then in equilibrium.

If now kinetic energy be in some way added to the constituent atoms of the group, atoms will be thrown out at a greater rate, and the density of the atmosphere must be increased to cause correspondingly greater condensation. Thus, the system remains in equilibrium at increasing temperature and pressure, and a stage will finally be reached at which the group may pass from one configuration of equilibrium to another. (Compare the two possible arrangements of closest homogeneous packing above mentioned.) If the process be now reversed, the group may remain in the new configuration. Kelvin believes that this indicates the process by which occur certain structural changes which have been shown by Madan to take place in chlorate of potash as its temperature is raised. Mallard has similarly explained other structural changes which are produced in crystals by variations of temperature or pressure.

If the process of adding energy to the constituents of the group, and increasing the density of the surrounding atmosphere, be carried further, a stage will ultimately be reached at which no configuration of equilibrium is possible, though the constituents still keep well within the range of their mutual attractive influence. If the external action be now stopped, we have a liquid and its vapour in equilibrium. If the process be carried further, the density of the vapour increases, while that of the group diminishes. At a certain stage, corresponding to the critical temperature, these densities become equal. Beyond that stage we have a homogeneous fluid at temperatures more and more remote from the critical point.

The kinetic theory must, in order to explain known phenomena of matter, be put in *some* Boscovichian form. Which form is the most suitable can, of course, only be determined by a contrast of experimental results with the results deduced theoretically from the assumption that that form is the correct one, or by finding out which law of force is necessitated for the explanation of certain facts. The vortex theory is a special type of Boscovichian theory in which the law of force is a foregone conclusion when the fundamental postulate has been made.

CHAPTER XII.

SOUND.

183. The word *sound* is generally used with reference to the physiological effect which results from excitation of the hearing organ. In physical science the term is applied to the external cause of this subjective impression.

We are accustomed to say that sound travels *through* a given medium, whether solid, liquid, or gas; and by this we imply that the particles of the medium do not move forward from the source of the sound to the place at which it is heard. Intermittent impact of such particles would account for the vibration of the tympanum which is necessary for the production of the mental impression of sound. But it is quite obvious that the particles of a solid cannot move forward in the manner indicated, and it would be unscientific to assume that the method of transference of sound in a gas differs totally from the method of transference in a solid body, especially when we consider that sound, after travelling through a solid, may be communicated to, and travel through, a gaseous medium, such as air. We do not, however, need to rely upon any such semi-metaphysical argument, for the speed of sound in air is so great that the forward motion of the particles—did it occur—would correspond to a hurricane of much greater violence than any ever observed. In short, we *know* that the passage of sound through air is not accompanied by such motion—witness the well-known fact that sound is usually best heard in *still air*.

But the passage of sound through air may communicate vibratory motion to objects immersed in it—*e.g.*, the tympanum—and it therefore involves transference of energy, which implies motion of matter. This motion can be nothing else than vibratory motion of the particles of the medium, the state of motion being handed on from particle to particle in the form of a wave which may cause vibratory motion of any solid object which it reaches, just as a wave sent along a stretched cord may cause motion of any object to which the cord is attached.

The vibrations of the particles of a medium may be transverse to the direction in which a wave travels, or they may take place in that direction. A little consideration will show that, in a sound-wave, it is the latter which must take place. Let us suppose that a sound is started by an explosion at a certain point. We know that the sound travels outwards in all directions from this point as centre; and we know also that the particles at the point are driven outwards by the explosion, so as to produce a state of great condensation in the immediate neighbourhood. But, the medium being elastic, the compressed portions expand, and so cause compression of the adjacent parts, and thus the state of compression travels outwards from the centre. Since a state of rarefaction is produced at the very centre of the explosion, the particles in the compressed part just outside rush back so as to fill the partial vacancy, and so the state of compression is succeeded by a state of rarefaction, which travels outwards at the same rate. Thus we see that sound consists in the propagation of a condensational-rarefactional wave, the particles of the medium vibrating to and fro in the direction in which the wave travels. It is easy to construct a model, which will roughly illustrate this process, by means of a row of equidistant balls which are attached by strings of equal lengths to a horizontal straight rod. Each ball must be set swinging through a small range in the plane of the whole system, the second ball being started a little later than the first, the third an equal amount later than the second, and so on. If, for example, the period of oscillation of each ball be one second, while each is started one-tenth of a second later than the preceding, the motions in each set of ten balls will be precisely repeated in the simultaneous motions of the next set of ten. F. J. Smith has described a simple method of performing the experiment. Displace all the balls, perpendicularly to their own plane, to an equal extent, by means of a board whose upper (straight) edge is not horizontal. If now the board be drawn downwards at a constant rate till all the balls are free, waves of transverse oscillation will be produced, and these may be changed into to-and-fro waves by simply turning the supporting bar horizontally through a right angle.

The distance, measured in the direction of propagation, from any particle to the next similarly-moving particle, is called the *wave-length*. It may be measured, for example, between points of maximum condensation, or between points of maximum rarefaction. The actual motion of any particle is a simple harmonic motion, and the terms used in the discussion of such motion (§ 38)—*amplitude*, *phase*, *period*, *etc.* — are therefore used here with similar meanings. Since the motion is simple harmonic, each particle has returned to its original position in the time in which the sound is propagated through one wave-length; so that we may define the wave-length as the distance through which the sound travels in the medium in the time in which the sounding body makes one complete vibration.

It is customary to speak of a non-periodic disturbance propagated through the air as a *noise*. The term *sound* is employed when a periodic disturbance, or series of disturbances, is propagated—the sound being *musical*, or *non-musical*, according as the disturbances are of regular or irregular period. When a musical sound is produced, it is usual to speak of the *note*, or *tone*, which is given out, but it is better to limit the application of the latter term to a simple sound of one definite period only. We shall see subsequently that no sound usually given out by a musical instrument consists of a simple tone only. The term *note* may conveniently be applied to the composite sound which is actually given out.

All sounds differ from each other in three points only—*intensity*, *pitch*, and *quality*. These will be treated in detail afterwards.

134. We shall now proceed to consider the propagation of sound through any gaseous medium. For the sake of simplicity, we shall suppose that a disturbance is propagated in the form of a plane wave; that is to say, we assume that any continuous set of particles, the motion of which is in the same phase, lies in a plane.

We have already seen (§ 56) that the square of the speed of a condensational wave is equal to the ratio which the resistance to compression bears to the density. If we assume that Boyle's Law holds, the resistance to compression, being the reciprocal of the compressibility, is numerically equal to the pressure (§ 80). Hence, p being the pressure and ρ the density, while v is the speed, we have

$$v^2=\frac{p}{\rho}=\mathrm{R}t,$$

where t is the absolute temperature; so that, in a gas which obeys Boyle's Law, the speed of sound is independent of the pressure, and is proportional to the square root of the absolute temperature. This expression is due to Newton. But, if the speed of sound in air at ordinary pressure and temperature be calculated by means of this formula, the calculated value will amount to about five-sixths of the true value only. This was observed by Newton, who accounted for the discrepancy by the hypothesis that it was due to the finite size of the particles of air. He assumed that one-sixth of the total space was occupied by the particles themselves, and that no time was occupied in propagation through them, so that the calculated speed must be less than the actual speed by one-sixth of the value of the latter. But Laplace pointed out that the compressions which take place when sound passes through a gas occur so rapidly that the heat which is developed by the work of compression has not time to spread outwards from the compressed parts, so that the temperature of these parts must be higher than the average temperature of the air; and the dilatations occur so rapidly that the fall of temperature which ensues (necessarily, as we see by the principle of stable equilibrium, § 15) cannot be compensated by a flow of heat into the dilated parts, so that the temperature of these parts must be lower than the average temperature. He showed that, in this case, the equation $p=\mathrm{R}t\rho$ does not accurately represent the actual conditions. Instead of it we must write $p=\gamma\mathrm{R}t\rho$, where γ is the ratio of the specific heat of the gas at constant pressure to its specific heat at constant volume. The speed is then given by the equation

$$v^2=\gamma\mathrm{R}t,$$

from which we see that the value of the speed as calculated by Newton's formula must be increased in the ratio of the square root of γ to unity. As the value of γ is 1·41 (§ 245), this ratio is practically 6/5, so that Laplace's result fully accounts for the discrepancy in explanation of which Newton introduced his erroneous hypothesis.

The following table shows how nearly the theoretical relative speeds correspond to the actual relative speeds as observed in a few of the more common gases:

	Observed.	Calculated.
Air	1·000	1·000
Carbonic Acid	0·786	0·811
Oxygen	0·953	0·952
Hydrogen	3·810	3·770

These numbers may be changed into absolute values in centimetres per second by the factor 33,240, which is the speed of sound in air at 0° C., as determined by Stone in 1871. Regnault's determination, in 1865, gave the number 33,104. The factor 1,090 reduces them to their values in feet per second.

135. The speed of sound in water is nearly four times as great as it is in air. This was found experimentally by Colladon and Sturm, in 1826, by means of observations made at the Lake of Geneva,

the waters of which are comparatively free from currents. A bell was sounded under water, and the mechanism which rang the bell simultaneously fired some gunpowder placed at the surface of the water at the same spot. A large receiver filled with air was placed below the surface at a known distance from the bell. An observer, who listened at the extremity of a tube connected with the receiver, knew at what instant the sound which travelled through the water reached the receiver. He also knew, by the flash of the gunpowder, the instant at which the sound was produced; and so he could calculate the speed at which the sound travelled through the water. The result was 1,435 metres per second at 8·1° C.

In solids the speed is still greater. Biot first showed this fact in the case of iron. A blow being struck by a hammer at one end of a great length of iron pipes, gave rise to two disturbances—one propagated through the air, the other through the material of the pipes. The latter was first heard at the far end of the set of pipes. Chladni, using a method to be described later (§ 148), gave (among others) as the values of the speeds in tin, silver, copper, iron, glass, oak, fir, relatively to the speed in air, the numbers 7·5, 9·0, 12·0, 16·7, 16·7, 10·7, 18 respectively. The factor 33,240 reduces these to absolute values in centimetres per second. Wertheim gives the value 501,600, in these units, as the speed in iron at 0° C. It is easy to calculate values of the speed in various substances, by means of the formula $v^2=k\rho$, from the data given in the tables in §§ 70, 89, 106.

The speed in air may be found by observing the interval of time which elapses between the instant of seeing the flash of a distant gun and the instant at which the report is heard. If the air is not still, two simultaneous observations must be made, in one of which the sound travels with the wind, and in the other of which it goes against the wind. The mean of the two results must be taken. It is well to remark here that the speed of sounds of explosive intensity may differ from that of sounds of ordinary intensity. Parry relates that, in the Arctic regions, the report of artillery was frequently heard before the preceding command of the officer to fire. Other methods of determining the speed in gases will be given afterwards (§§ 139, 147).

136. The loudness, or *intensity*, of a sound is due to the kinetic energy per unit volume of the medium, and so it depends upon the square of the speed of oscillation of the particles. Now, the average speed of oscillation is necessarily proportional to the amplitude if the period be fixed. Hence the intensity is proportional to the square of the amplitude. It depends also, other things being equal, upon the density of the medium, being greater as the density is greater,

and vanishing when the density becomes zero. Thus, no sound is heard when a bell is struck inside a receiver from which the air has been exhausted as completely as possible.

In still homogeneous air sound spreads outwards from the source uniformly in all directions; and therefore particles whose vibrations are in a given phase lie on the surface of a sphere, the radius of which grows uniformly. Since the energy of vibration is distributed over the surface of this growing sphere, it follows that the intensity diminishes proportionately as the surface increases. It is therefore inversely proportional to the square of the distance from the source.

Sound travels faster than usual when the air through which it moves is blowing in the direction in which the sound moves, for the two speeds are simply superposed. Similarly, when the wind is blowing in a direction opposite to that in which the sound moves, the speed is distinctly diminished. And, in addition to this, motion of the medium affects the distance at which the sound may be audible. For, if *ab* represent the (vertical) front of a plane-wave of sound travelling with the wind in the direction *ax*, it is evident, since the motion of the upper strata of air is less retarded by friction than that of the lower strata, that the wave-front gradually becomes more and more inclined to the vertical as it moves forward. Successive new positions are indicated by the lines $a'b'$, etc. And the sound, instead of travelling straight forwards, is thrown down towards the ground in the manner indicated by the curved lines in Fig. 56. When the wind blows in a direction opposite to the direction of motion of the sound, the sound is thrown up from the earth's surface so as to be inaudible at comparatively short distances. These results were pointed out by Stokes.

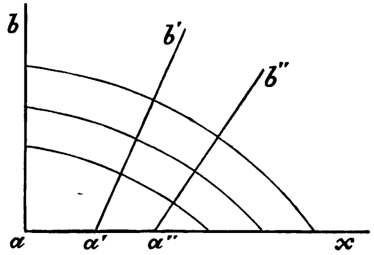


FIG. 56.

Osborne Reynolds remarks that refraction (§ 138) produces similar effects. If the density decreases from the surface of the earth upwards, the sound will be thrown downwards.

137. *Reflection of Sound.*—When sound strikes an obstacle, it is reflected in such a way that *the reflected ray* (the word is used by analogy from the phenomena of light, which, we shall see subsequently, are due also to wave-propagation) *and the incident ray make equal angles with, and lie in a plane passing through, the normal to the surface.* Thus, if *ab* (Fig. 57) represent a plane surface from which the ray *ec* is reflected in the direction *cf*, the line *cf* lies in a plane passing through *ec* and the normal *cd*, and the angles *ecd* and *fcd* are equal.

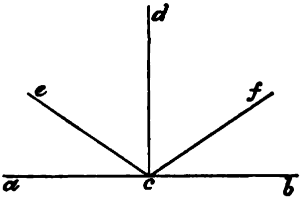


FIG. 57.

The law is identical with that which obtains in the reflection of light, so that all the results which are deduced in Chap. XVI. regarding the reflection of light from surfaces will at once apply to the case of reflection of sound. It may be proved experimentally by the employment of a *sensitive flame.* Such a flame is fed by gas at a high pressure. If the pressure be too high, the flame will flare. If the pressure be slightly below the value required to make the flame flare, flaring will be started when the air in the neighbourhood is set in vibration by means of sound which may even, because of its high frequency of vibration (§ 141), be inaudible to the ear. If the flame be set in a position in which it is most sensitive to the sound, it will cease to be sensitive if the burner is turned in either direction through a right angle about its vertical axis. A further rotation of the burner through a right angle again makes the flame sensitive. By placing the burner in the proper position, or by otherwise screening it, so that the flame does not respond to the direct sound, we may find the positions in which it is most sensitive to a reflected sound, and so prove the law of reflection. Care must be taken to have the reflecting surface, and any screen that may be employed, large in comparison with the wave-length of the sound, so that diffraction (§ 140) does not interfere with the result. The law can be deduced as a result of the fact that sound consists in wave-motion, in precisely the same way as that in which the corresponding law is deduced in § 161 as a result of the wave-theory of light.

Echoes are due to the reflection of sound from buildings, rocks, trees, clouds, etc. The so-called 'rolling' of thunder is due to reflection from clouds. Reflections occur even where there is no visible object to account for their existence. In this case they must originate at the common surface of two large masses of air of different densities, or containing very different amounts of moisture per unit volume. Thus, Tyndall has demonstrated the reflection of sound from a mass of heated air. The ear cannot tell that two sounds are not simultaneous if they succeed each other too rapidly. Hence the existence of an echo cannot be recognised if the observer is within a certain distance of the reflecting surface. If the reflecting surface be curved, the reflected sound may be conveyed to a focus so as to be much more distinctly heard than the direct sound, as in the whispering-galleries of St. Paul's and the Colosseum.

138. *Refraction of Sound.*—Since the speed of sound has different values in media of different densities, and since, in any one homogeneous medium, the sound spreads out uniformly in all directions from a centre of disturbance, it follows that the direction in which a ray is travelling is, in general, suddenly changed when the sound passes from one such medium into another. If it pass from a less dense into a more dense medium, the direction of propagation is inclined to the normal at a smaller angle after the interface is passed than before; and the opposite statement holds when the first medium is denser than the second. This phenomenon is known as the refraction of sound. The law is that *the incident and the refracted rays lie in one plane with the normal to the refracting surface, and make with it angles whose sines bear a*

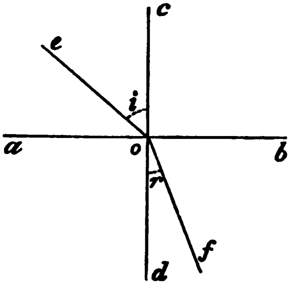


FIG. 58.

constant ratio to each other. If *ab* (Fig. 58) represent the intersection of the common surface of the two media by the plane of the paper, while *cd* is the normal, and if *eo* be the direction in which the sound impinges upon the surface, while *of* is the direction which it takes after entering the second medium, the angles $i = eoc$ and $r = fod$ are connected by the relation $\sin i = \mu \sin r$, where μ is a constant.

These results are identical with those which are observed in the refraction of light (§ 162), and the reasoning of § 175 may be applied directly to the present case. A lenticular bag, filled with carbonic acid gas, has been found to convey sound to a focus in precisely the same manner that a glass lens conveys light to a focus.

139. *Interference of Sound.*—As sound consists physically of waves of condensation and rarefaction, it follows that two sounds may 'interfere' with each other, according to the usual laws of interference of waves. Thus, if two sounds are travelling in the same direction through a given medium, and if the intensity and wave-length of these sounds are identical, there will be no resultant disturbance of the medium, *i.e.*, no sound will be heard, provided that the maximum condensation due to one of the disturbances occurs simultaneously with the maximum rarefaction due to the other. If both condensations, and therefore both rarefactions, take place together, a sound of four times the intensity will be heard, for the resultant amplitude of vibration of the particles of the medium is doubled. And, if the wave-lengths of the separate disturbances are not precisely identical, the resultant sound will periodically vary in intensity from the former (zero) to the latter (quadruple) value.

Some of these phenomena are very readily shown by means of an apparatus in which sound is sent along a tube which is symmetrically divided, throughout part of its length, into two branches. When the sound reaches the point of division, one half of it goes along the one branch, and the other half goes along the other branch. These separate sounds combine again at the point where the branches of the tube reunite, and proceed as a single sound along the remainder of the tube. Under the condition of complete symmetry of the two branches which we have postulated, the resultant sound is precisely identical with the original sound—except in so far as there may be a slight diminution of intensity by friction of the air against the tube, or by communication of vibrational energy to the walls of the tube, which loss we neglect. But if the one branch differs in length from the other, a difference of phase will be produced between the components of the resultant sound; and this difference of phase will become half a period when the difference of length is half a wave-length. Under the latter condition no sound will be heard by an ear placed at the end of the tube; and, by measuring the difference of the lengths of the two branches, we can determine the wave-length of the sound. It is essential, of course, that the initial sound should have a definite wave-length. This may be attained by using a tuning-fork to originate the sound, and by attaching a resonator (§ 148), tuned to the fundamental tone of the fork, to the end of the tube at which the sound enters.

It is easy to make the effect evident to the eye. For this purpose the resultant sound may be conveyed into one compartment of a receiver, which is divided into two portions by a flexible diaphragm, and whose other compartment contains ordinary coal-gas, at constant pressure, which supplies a lighted jet. If the two components of the sound are in opposite phases, the diaphragm remains steady, and so no change is caused in the jet. If the components are in like phases, the resultant sound makes the diaphragm vibrate, and so causes variations of pressure in the compartment containing the coal-gas. This, in turn, causes the jet to flicker. The motion of the jet cannot be directly seen by the eye; for the eye cannot distinguish two flashes which occur with an interval of less than about one-tenth of a second from a single flash, and the vibrations of all audible sounds have greater frequency than this. But, by viewing the flame in a rotating mirror, the variations in the jet can be readily seen.

A sensitive flame (§ 137) might be used to show when the resultant sound vanishes.

It may be fancied that, in the above apparatus, there is a destruction of energy. We have assumed that waves travelling along the two branches proceed together along the single tube where the branches recombine; and we have found that, in one case, they do so in such a way that there is no resultant disturbance. Energy constantly enters the tube at one end, no energy leaves it at the other, and there is never more than a finite amount inside it. It should scarcely be necessary to point out that the conclusion that two mutually destructive waves pass along the tube is simply equivalent to the conclusion that no waves pass, and we therefore infer that they go elsewhere. Each wave passes back along one or other of the branches, giving rise to a 'stationary wave' (§ 85), and the energy leaves the tube by the end through which it entered.

The case of stationary waves is of special interest, and will be further discussed when we consider the vibrations of rods, strings, and air-columns.

140. *Diffraction of Sound.*—Sound which enters a room by an

aperture such as a window, is equally well heard at all parts inside the room. It bends round so as to penetrate every portion, and does not cast a sharp 'shadow' of an obstacle as light does. This bending of sound into the region behind an obstacle is termed *diffraction*. We shall see afterwards that the same phenomenon also occurs, under suitable conditions, in the case of light.

But it is possible by proper means to produce sound shadows. The necessary condition is that the aperture through which the sound passes, or the obstacle by which it is intercepted, shall be large in comparison with the wave-length of the sound. Thus a sound made at one side of a steep high bank may be totally unheard at the other. So, also, a sound may be clearly heard through a hollow between two mountains, while it is inaudible in the region behind either mountain. And, indeed, by using sound of sufficiently short wave-length, we can produce comparatively sharp sound shadows of obstacles which are only a few centimetres in diameter.

With obvious verbal changes (such as replacing 'light' by 'sound,' etc.), and apart from historical statements, the whole of §§ 198-204 may be read in this connection. All the phenomena described in these sections can be demonstrated in the case of sound by the use of sensitive flames and sound-waves from 1 to 2 centimetres in length. Rayleigh has thus demonstrated some of them to a large audience.

141. *Pitch.*—The resultant vibration of a particle of air is, in general, extremely complex; but, when a pure tone is transmitted through air, the vibration of each particle is simple harmonic. There can, therefore, be no difference between one tone and another except such as is due to a difference in the amplitude of vibration or to a difference in the period of vibration. As the amplitude of vibration determines the intensity of the sound, we infer that the pitch of a note depends upon the frequency of vibration, *i.e.*, upon the number of vibrations which are executed per second.

That this is actually the case may be roughly ascertained by means of very simple apparatus. If a piece of cardboard be pressed against the edge of a toothed wheel, which revolves at a definite rate, a sound, which is of a fairly definite pitch, is emitted. As the speed of the wheel increases, *i.e.*, as the number of impacts per second between the teeth of the wheel and the cardboard increases, the pitch of the note rises; and, to a note of given pitch, there corresponds a definite rate of rotation of the wheel, and consequently a definite rate of vibration of the cardboard. This arrangement is known as Savart's toothed wheel.

A much more accurate proof is obtained by means of the *siren*.

Seebeck's form of this instrument consists essentially of a perforated disc of metal, the perforations being arranged in a circle round the centre, as in the figure. A tube, through which air is driven, is placed behind the disc; and, as the disc revolves, each successive opening in it comes opposite the end of the tube. As each blast of air passes through, a state of condensation is produced, which is succeeded by a rarefaction in the interval between two blasts. Hence a sound-wave of constant period is set up when the disc revolves at a uniform rate. The rate of rotation, and the number of perforations in the complete circle, being known, we get at once the number of vibrations per second corresponding to any note of given pitch. In Cagniard de la Tour's form of this apparatus the holes are drilled obliquely through the disc, which is consequently made to rotate automatically by the impact of the air; and the further improvement of placing a correspondingly bored fixed disc behind the rotating disc is employed, the inclination of the holes being opposite to that of the holes in the revolving disc. The fixed disc forms the end of a compartment through which the air is blown. The air, therefore, passes through all the holes simultaneously, and so the intensity of the sound is increased. In modifications of this siren more than one concentric series of openings are used, any or all of which can be employed at will. In Von Helmholtz's double siren the (usually) fixed disc of one part can be slowly revolved at known rates, so as to produce variations in the otherwise fixed relative pitch of notes which are being simultaneously sounded.

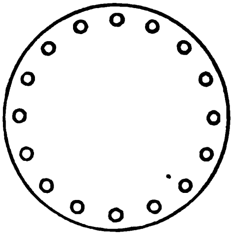


FIG. 59.

When the disc revolves very slowly, no musical sound is heard, but each separate air-blast can be distinctly heard. As the speed increases, the ear ceases to distinguish the separate pulses, and a note of very low pitch becomes audible. The speed still increasing, the pitch of the note produced becomes higher and higher, and at last it becomes so high that the note is no longer audible. The limits of audibility vary considerably in different observers, but, roughly speaking, a sound becomes imperceptible as a continuous note when the rate of vibration falls short of 20 times or exceeds 40,000 times per second. The fact that music is perfectly heard at different distances from the source shows that the speed of sound does not depend upon the wave-length.

The apparent pitch of a note depends upon the relative motion of the hearer and the instrument upon which the note is sounded. If the hearer approaches the instrument, the pitch of the note seems to be heightened, because more than the normal number of vibrations reach his ear in a given time; and, conversely, if he recede from the instrument, the pitch of the note will appear to be lowered. As a particular case, if the rate of recession were greater than the speed of sound, the instrument and the air being at rest, no sound could be heard. An unsteady wind would produce variations of pitch, though a steady wind would have no effect.

142. *Musical Intervals.*—If any particular note be taken as a keynote, it is found that there are certain other notes, definitely related to it as regards pitch, which produce specially pleasing effects upon the ear when sounded with each other or with the keynote. The reason for this will appear in § 150. These are, therefore, the notes which are employed in the ordinary major and minor scales, and the difference in pitch of any two notes is called the *interval* between them. When two notes have the same pitch, they are said to be in *unison*.

The chief interval is the *octave*, and it is found that, in order to produce the octave of any given note, the rate of vibration must be exactly doubled. The intervals into which the octave is subdivided are not equal. The following table indicates their values. The first column gives the names of the intervals which separate each note from the keynote, and the relative rates of vibration in the two notes forming each interval are given in the second column. For example, nine vibrations are performed in the higher of two notes separated by the interval of a second in the time in which eight vibrations are performed in the lower:

Unison	$\frac{1}{1}$.
Second	$\frac{9}{8}$.
Minor Third	$\frac{6}{5}$.
Major Third	$\frac{5}{4}$.
Fourth	$\frac{4}{3}$.
Fifth	$\frac{3}{2}$ or $2(\frac{3}{4})$.
Minor Sixth	$\frac{8}{5}$ or $2(\frac{4}{5})$.

Major Sixth	$\frac{5}{3}$ or $2(\frac{5}{6})$.
Minor Seventh	$\frac{16}{9}$ or $2(\frac{8}{9})$.
Major Seventh	$\frac{15}{8}$.
Octave	$\frac{2}{1}$.

A very little consideration will show that in order to find the sum of two intervals we must multiply together the fractions given above for each interval. Thus, if three notes, *a*, *b*, and *c*, be in ascending order as regards pitch, the interval between *b* and *a* being a minor third while that between *b* and *c* is a major third, *c* makes 30 vibrations in the time in which *b* makes 24, and *b* makes 24 in the time in which *a* makes 20. Thus the vibration-frequencies of *c* and *a* are in the ratio of 3 to 2, so that the interval between these notes is a fifth. Thus a fifth is the sum of a minor third and a major third. Also, to find the difference between two intervals we must divide the fraction corresponding to the larger by the fraction corresponding to the smaller. The second method of writing the fractions corresponding to the minor seventh, the major and minor sixths, and the fifth, shows that the interval between the notes so indicated and the octave are respectively a second, a minor third, a major third, and a fourth.

143. *Vibrations of Rods.*—We have now to consider the laws of the production of notes by sounding bodies. We shall assume that the extent of the vibrations is such that Hooke's law (§ 111) is followed. In this case, since the period of vibration is independent of the extent of vibration, a musical note of constant pitch will be heard, provided that the rate of vibration is sufficiently rapid. Two kinds of vibration have to be considered.

Transverse Vibrations.—We see from §§ 33, 46 that the time of simple harmonic vibration of any material system varies directly as the square root of the mass to be moved and inversely as the square root of the stress called into play by a given displacement. For the square of the time is proportional to the ratio of the displacement to the acceleration, while the stress is equal to the product of the mass into the acceleration. Again, the stress called into play by a given displacement is proportional to the flexural rigidity and the displacement. Hence the time, in the case under consideration, varies inversely as the square root of the flexural rigidity. Now both the flexural rigidity and the mass of a rod are (§ 109) proportional to its breadth, and so *the period of transverse vibration is independent of the breadth of the rod.* Again, the mass of the rod is proportional to the length, while the rigidity varies inversely as the cube of the length. Hence the period is *proportional to the square of the length.* Similarly, *it varies inversely as the thickness,* last it becomes so high that the note is no longer audible. The limits of audibility vary considerably in different observers, but, roughly speaking, a sound becomes imperceptible as a continuous note when the rate of vibration falls short of 20 times or exceeds 40,000 times per second. The fact that music is perfectly heard at different distances from the source shows that the speed of sound does not depend upon the wave-length.

The apparent pitch of a note depends upon the relative motion of the hearer and the instrument upon which the note is sounded. If the hearer approaches the instrument, the pitch of the note seems to be heightened, because more than the normal number of vibrations reach his ear in a given time; and, conversely, if he recede from the instrument, the pitch of the note will appear to be lowered. As a particular case, if the rate of recession were greater than the speed of sound, the instrument and the air being at rest, no sound could be heard. An unsteady wind would produce variations of pitch, though a steady wind would have no effect.

142. *Musical Intervals.*—If any particular note be taken as a keynote, it is found that there are certain other notes, definitely related to it as regards pitch, which produce specially pleasing effects upon the ear when sounded with each other or with the keynote. The reason for this will appear in § 150. These are, therefore, the notes which are employed in the ordinary major and minor scales, and t[illegible]ference in pitch of any two notes is called the *interval* be[illegible]her[illegible]n two notes have the same pitch, they are [illegible]

The [illegible]in[illegible]

produ[illegible]*ve*, and it is found that, in order to exact[illegible] note, the rate of vibration must be are[illegible]into which the octave is subdivided [illegible]indicates their values. The [illegible]ervals which separate each [illegible] rates of vibration in the [illegible]n in the second column. [illegible]med in the higher of two [illegible]ond in the time in which [illegible]er:

[illegible] ...	$\frac{1}{1}$.
[illegible] ...	$\frac{9}{8}$.
[illegible] ...	$\frac{6}{5}$.
[illegible] ...	$\frac{5}{4}$.
[illegible] ...	$\frac{4}{3}$.
[illegible] ...	$\frac{3}{2}$ or $2(\frac{3}{4})$.
[illegible] ...	[illegible] or $2(\frac{[illegible]}{3})$.

sectional area. An experimental method of determining the speed will be given in § 147.

The rod which is fixed at both ends has twice the rate of vibration of a similar rod fixed at one end only. A rod which is free at both ends, and has its node at the centre, obviously vibrates at the same rate as does a like rod with both ends fixed.

144. A rod, which is fixed at one end, or at both ends, or has its ends free, has more than one mode of transverse or longitudinal vibration. Certain points, the nodes, remain at rest, while the portions of the rod between adjacent nodes form loops in vibratory motion. The motions of adjacent loops are at every instant oppositely directed. In fact, the state of motion of the rod is that of a 'stationary wave,' due to the simultaneous existence of two equal waves moving in opposite directions with equal speeds, each wave being reflected when it reaches an end of the rod, whether fixed or free.

The case of longitudinal vibrations is the simpler. The modes of vibration are identical with those of air-columns, a free end of the rod corresponding to an open end of an air-column, and a fixed end of the rod corresponding to a closed end of an air-column; so that § 147 may be read in this connection with obvious verbal changes.

The possible modes of transverse vibration are more complicated. The periods are the same for a rod fixed at both ends as for a rod free at both ends. In the latter case the fundamental mode has a central loop and two free end parts. The next higher vibration has two loops and two end parts, and so on. The rule giving the number of vibrations per second for the fundamental and higher forms may be stated in the following way: Take rods of different lengths but otherwise alike, the lengths being proportional to the odd natural numbers, not including unity. Set them all vibrating transversely in any manner whatsoever, provided only that it is the same for them all. The periods of these, in ascending order, will be inversely proportional to the periods of the fundamental and higher transverse vibrations of a bar fixed, or free, at both ends. As we have found that the period of vibration of a bar, vibrating in a given mode, is proportional to the square of its length, we see that the fundamental and higher vibrations of a given bar, fixed or free at both its ends, are inversely proportional to the squares of the odd natural numbers, excluding unity. The number of vibrations per second is the reciprocal of the period, and is called the *vibration-frequency*. The frequency of the n^{th} higher vibration, relatively to the frequency of the fundamental vibration, is therefore given by the expression $(2n+3)^2$.

In the case of the rod fixed at both ends, it might be imagined that the frequencies should proceed as the squares of the complete series of natural numbers since the number of loops proceed as the natural numbers. It must be remembered that the loops are not all alike, as they are in the vibrating string (§ 146). In the case of a string, the fixed end acts as a true node at which the little element of the string rotates backwards and forwards through a finite angle once in each half-vibration. In the case of the rod this rotation is prevented at a fixed end though it takes place at the intermediate nodes.

In the case of a rod fixed at one end, which is practically the condition in each limb of a tuning-fork, the frequencies of the fundamental and the next higher notes are proportional to 2^2 and 5^2 respectively.

145. *Vibration of Plates.*—A plate may be set in vibration by drawing a bow across its edge. The modes of vibration depend on the form of the plate, and may be greatly varied by forcing certain points to lie upon nodal lines—which may be done by keeping the fingers in contact with the desired points. Chladni exhibited the nodal lines by strewing sand upon the plate. When the vibrations are started the particles of sand are driven away from the moving parts and collect along the nodal lines. Lycopodium powder, if used instead of sand, is carried away from the nodal lines and collects at the regions of greatest vibration. Faraday showed that this effect vanishes in a vacuum. In accordance with the principle that in a fluid the pressure is least where the speed is greatest, air-currents flow in towards these regions carrying the light powder with them.

As a special example, the nodal lines of a square plate, clamped at its centre and touched by the finger at one corner while the bow is drawn across the middle of the edge, are the diagonals. The plate is, of course, an extreme case of a rod, and has the same modes of vibration as a rod if its fixed lines or nodal lines are parallel to one of the pairs of edges. Wheatstone showed that even the most complicated forms of vibration of a plate result from the superposition of its vibrations with respect to two mutually perpendicular directions in its own plane.

In the above example we may suppose that the plate is free, and that it vibrates in its fundamental manner so as to have two nodal lines parallel to one of its pairs of edges. Suppose that the central part is raised while the edges descend. Let it vibrate also with respect to its other pair of edges, so that the edges are raised while the central part descends. The superposition of these two fundamental rod-vibrations gives the resultant plate-vibration with its diagonal nodal lines.

If the plate is circular with its centre clamped, the nodal lines are diameters which divide the plate into equal sectors—necessarily even in number, four being the fundamental number. The higher tones have, therefore, frequencies proportional to the squares of the successive even numbers from four upwards. If the circular plate be given curvature symmetrically on all sides of its central point, an approximation is made to the case of a bell.

As in § 143, we can without difficulty deduce the result that the periods of vibration of similar plates (similar, of course, as regards thickness as well as form) of the same material are in proportion to their linear dimensions. Also, as in the section referred to, we may show that the period is inversely as the thickness of a plate of given form and area. By combining these two results we see that the periods of vibration of similar plates of the same thickness are proportional to the squares of their linear dimension, that is, to their areas.

146. *Vibrations of Strings and Membranes.*—The speed, v, with which a disturbance runs along a stretched cord is (§ 56) proportional to the square root of the tension, T, and is inversely proportional to the square root of μ, the mass per unit length of the string. Hence, if λ be the wave-length, the time of a complete vibration is

$$\frac{\lambda}{v}=\lambda\sqrt{\frac{\mu}{\mathrm{T}}}=\lambda\sqrt{\frac{\rho s}{\mathrm{T}}},$$

where ρ is the density of the material of the string and s is its area of cross-section.

The figure below shows the three simplest forms of vibration of a stretched string which is fixed at its ends. Any disturbance of the string is propagated with equal speed in both directions along the string from the point of disturbance, each part being reflected when it arrives at a fixed end. Hence (§ 35) the string is thrown into segments, the wave-length in each case being equal to twice the length of a loop.

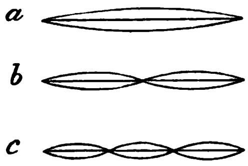


FIG. 60.

Thus the above equation enables us to assert that *the fundamental period of vibration of a string is directly proportional to its length, to the square root of its density, and to the square root of its sectional area, and is inversely proportional to the square root of its tension.*

If a string of length l vibrates n times per second (determined by means of the siren), the speed of the wave is $2ln$.

The higher vibrations correspond to tones which are respectively one, one-and-a-half, two, two-and-a-quarter, etc., octaves above the fundamental tone. It is interesting to compare this result with the corresponding result in the case of a vibrating rod which is fixed at both ends. In the latter case, the transversely propagating force is the flexural stress; in the present case, it is the tension of the cord, in which flexural stress has no existence.

Stretched membranes bear the same relation to stretched strings as plates do to rods. They can vibrate independently with regard to two rectangular directions in their own plane.

147. *Vibration of Air-Columns.*—When a condensation reaches the closed end of a pipe, or tube, containing air through which sound is travelling, the increase of pressure which takes place can only be relieved by expansion of the air backwards along the pipe. Similarly, when a rarefaction reaches the closed end, the consequent diminution of pressure can only be changed by a flow of the air contained in the pipe towards that end. In other words, the wave is simply reflected at the closed end, the result being that two sound waves are travelling simultaneously along the pipe, in opposite directions, with equal speed. And the state of pressure due to each is always in the same phase at the closed end of the pipe. But, at a certain distance from the closed end, an outgoing condensation meets an incoming rarefaction. Hence, on the assumption that no energy has been lost in the process of reflection, the pressure has here its normal value. And this normal condition is maintained constantly so long as the wave-length of the disturbance is unaltered; for the two oppositely-travelling waves are always in opposite phases as regards pressure at this place, which is evidently distant by one-quarter of a wave-length from the closed end of the pipe.

The conditions of motion of the particles of air are simply reversed when the wave is reflected, and, consequently, the incident and the reflected disturbances are in opposite phases as regards vibration when they are in the same phase as regards pressure or density. Hence the places of maximum variation of pressure are places of no vibration, and the places of uniformly normal pressure are those of maximum vibration. Thus a node occurs at the closed end of the pipe, and other nodes appear at equal intervals of half a wave-length.

It is obvious that a loop, or place of greatest motion, must occur at the open end of a pipe. For, when a condensation reaches it, the air is free to expand outwards—much more free to do so, in fact, than to expand inwards—and consequently the condensation is succeeded by a rarefaction which is propagated back through the pipe. Similarly, an incident rarefaction is filled up by the influx of the surrounding air, and so a condensation is returned. Thus we see that there is great vibration at the open end of the pipe, which therefore corresponds to a loop. This might otherwise have been determined by means of the consideration that, in the immediate neighbourhood of the open end, the air-pressure is uniform and equal to the normal atmospheric pressure. Consequently the open end is a place of maximum vibration.

Stopped and Open Pipes.—A pipe which is closed at one end is called a 'stopped' pipe, while one which is open at both ends is called an 'open' pipe.

In a stopped pipe the nodes evidently occur at the same positions as do the nodes of a transversely-vibrating string which is fixed at one end. This is indicated in the figures below, in which the vertical distance between the waved lines at any part of the length of the pipe may be supposed to indicate the extent of the vibrations at that part. The first figure shows the fundamental mode of vibration, The wave-lengths, and consequently the periods of vibration, in all the possible modes are evidently inversely as the odd numbers 1, 3, 5, etc.; and, in the fundamental mode of vibration, the length of the wave is four times the length of the pipe; for a condensation which leaves the open end has its phase reversed when it again

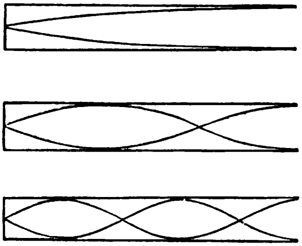


FIG. 61.

reaches that end and is reflected. Two reversals are therefore necessary in order that the original phase may be attained.

In an open pipe the nodes occur at the same positions as do the nodes of a transversely-vibrating string which is not fixed at either end. The positions are indicated below. The possible periods of vibration are inversely as the natural numbers 1, 2, 3, etc. The wave-length in the fundamental mode of vibration is twice the length of the pipe; for a condensation passing from one end of the pipe is reflected from the other end as a rarefaction, and leaves the original end once more as a condensation. Hence an open organ-pipe which has the same fundamental tone as a given closed organ-pipe must be twice as long as the closed one is.

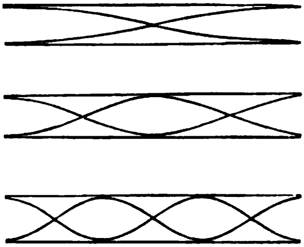


FIG. 62.

Pipes are divided into two classes: *flue-pipes*, such as the flute and the flageolet; and *reed-pipes*, such as the cornet, the oboe, the trumpet, and the trombone. In a flue-pipe, the air, passing through a narrow slit, impinges on a sharp edge and is deflected into the pipe, sending a wave of compression up the tube. After reflection at the end of the tube this wave meets a second out-going wave, and so the stationary vibrations are produced. In a reed-pipe, the *reed*—a thin strip of metal fastened at one end—is placed over the opening at which the air enters the pipe. If it can entirely cover the opening, it is called a *beating* reed; if it can vibrate through the opening, it is called a *free* reed. Air rushing past the reed sets it in vibration, and it is this vibration of the reed which determines the note sounded by the pipe. The air-column must be made of such a length as to have the fundamental period requisite (see § 148) for strengthening the main tone of the reed.

Speed of Sound and Measurement of Wave-length.—In order to determine the speed of sound in any given gas we merely require to sound an organ-pipe of given length which is filled with that gas. By means of the siren the number of vibrations which are made per second is found. And if l be the length of the pipe, while n is the number of vibrations made per second, the speed is $4nl$, or $2nl$, according as the pipe is closed or open. Friction of the gas against the sides of the pipe makes the result somewhat inaccurate, and more so the narrower the pipe is.

In one method, due to Kundt, a glass rod, about 1·5 metres in length and over 1 centimetre in diameter, has its ends corked after some lycopodium powder or (preferably) precipitated silica has been placed inside. The tube gives out a musical note if it be held lightly by the middle and be rubbed longitudinally by a wet cloth, and the silica is carried by the air-currents to the central parts of the loops. Twice the distance from one such centre to the next centre is the wave-length, and the vibration-frequency may be found by means of the siren. The product of these quantities is the speed of sound in the particular gas with which the tube is filled.

Kundt employed this method also to find the wave-length of any given note, a wider tube being used. This tube was closed, at one end only, by a movable piston, and the open end was presented to the source of sound. The position of the piston was varied until the air-column gave out the same note as the source (§ 148), at which stage the silica acted as in the above case.

If the end of a solid rod, which is made to give out its fundamental note by longitudinal vibration, be held just within the open end of the tube (with a guard if necessary, so as to completely fill the end of the tube) while the length of the air-column is altered as before, the speed of sound in the solid rod may be determined relatively to the speed in air. For the silica indicates the wave-length in air, while (§ 143) the wave-length in the solid is double the length of the rod; and the ratio of the wave-lengths is the ratio of the speeds.

148. *Partial Tones.—Resonance.*—From the results of the last few sections it is now evident that no note given out by a musical instrument is, strictly speaking, a pure tone; in most cases it very obviously is not so. Thus in instruments, such as the harmonium, in which reeds are employed without pipes, each metal strip has a very complicated vibration—a large number of the higher vibrations referred to in § 144 being strongly present as well as the fundamental vibration. In this way a highly complex note, in which occur tones corresponding to the various forms of vibration which the reed assumes, is produced.

These constituent tones of a note are called the *partial-tones;* and it is usual to distinguish between the *fundamental tones* (corresponding to the fundamental mode of vibration of the sounding body) and the higher partial-tones, which are termed the *overtones*.

We have already seen that the intensity of a sound depends, other things being equal, upon the amount of energy which is communicated to the air in each vibration of the body which originates the sound; and we saw also that it is proportional to the square of the amplitude of vibration. Hence energy must be supplied at a greater and greater rate to a body, vibrating successively to a given extent in different manners, in proportion as its vibrations are more and more rapid. For this reason alone we see that it is more difficult to produce an overtone than to produce the fundamental tone, if the amplitude, and therefore the intensity, is to be the same for each. Consequently the overtones of a note are in general feebler than the fundamental tone—each overtone, taken in order of pitch, being weaker than the preceding one. In this way a note generally has the same pitch as its fundamental tone, though it is, of course, possible so to act upon a sounding body that an overtone shall have greater intensity than the tone which corresponds to the fundamental vibration.

The unaided ear is able to detect in great measure the various overtones which are present in a given note; but, in this analysis, it may be greatly aided by means of instruments, the action of which depends upon the principle of *resonance*, according to which *any sounding body can readily absorb, and give out again of itself, a sound which is emitted by another sounding body which has a period of vibration identical with its own.*

To understand this principle, we need only refer to a well-known dynamical analogy: A pendulum of given length has a definite period of vibration, and oscillations of great magnitude may be induced in it by the application to the bob of the *feeblest* impulses, provided only that these impulses are communicated regularly at instants the interval between which is equal to the natural period of oscillation of the pendulum. The effects of all the impulses are in the same direction, and so the total effect may be very large although each single effect is excessively small. And, similarly, the feeble periodic impulses which are communicated through the medium of the air to a body which is capable of giving out sound may, by being properly timed, set that body in such a state of vibration as to give out the sound of itself after the original note has ceased.

The *resonator* of Von Helmholtz consisted, in its earlier form, of a hollow brass ball with two apertures at opposite ends of a diameter. This is indicated in Fig. 63. Sound is communicated to the air in the ball through the large aperture, and the small aperture is applied to the ear. The air inside the ball has a definite fundamental period of vibration, and therefore that tone (should such exist) which corresponds to this period is very distinctly heard—

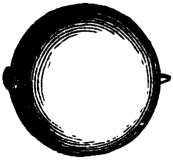


FIG. 63.

all others being entirely, or almost entirely, suppressed. The later form of the instrument is cylindrical. Mayer uses tuning-forks, instead of air-columns, as resonators for the analysis of sounds.

In many cases it is desirable to intensify the fundamental tone of a note. The principle of resonance shows that this may be done by associating with the sounding body a resonating column of air which has the same fundamental period of vibration. Thus a reed is fitted at one end of an open pipe of the proper length; and a tuning-fork is attached to a 'sounding-box,' which is simply a closed pipe of such a length that the contained air-column has the same fundamental period of vibration as the tuning-fork has. The box itself is set in vibration by the fork, and so the motion is communicated to the enclosed air.

It is essential to notice that the *box* to which the fork is attached is not itself set in vibration by resonance. It would act equally well with any fork, though the air-column inside it would not do so. The action of the sounding-board of a pianoforte is similar. The vibrations of the fork or wire are directly communicated to a body of large surface, and so are enabled to set a large mass of air in vibration. On the other hand, a vibrating wire, if attached to very rigid supports, would give out a sound of very feeble intensity, and its motion would last for comparatively a long time. The action of the box or of the sounding-board exemplifies the general principle that an elastic body, whose natural periods of vibration are sufficiently damped, will respond equally well—we might say equally badly—to vibrations of almost all periods.

149. *Quality.*—There is between two pure tones of the same pitch no difference of quality such as that which distinguishes the same note when played on two different instruments. Von Helmholtz, to whom are due the great developments of the whole subject,

therefore concluded that difference of quality between two notes is due to the existence of partial-tones.

Experiment shows that this is actually the case, and it indicates further that the quality of a note depends upon the number of partial-tones which are present in it; that, the number being the same, it depends upon the particular set of tones which go to make up that number; and, also, that it depends upon the relative intensities of the partial-tones.

According to Von Helmholtz, it is independent of the particular phase in which any one of the constituent vibrations may be when the phases of the others are given. Now, the nature of the resultant vibration essentially depends on this condition—§ 84 (1)—so that the assertion is equivalent to a statement that the quality of a note depends on the nature of its constituents alone. Kœnig denies this, and has devised an apparatus in which the same constituents are blended in different phases and produce different effects as to quality.

As a special example, we may instance the difference of quality between two notes of the same pitch sounded respectively on a closed and an open organ-pipe. In a closed pipe the odd partial-tones alone occur, while in the open pipe both the odd and the even tones are present.

Since any note may be resolved into its constituents by means of resonating tuning-forks or air-columns, it is possible, conversely, to combine constituent notes so as to produce a note of any required quality. In one apparatus Von Helmholtz used both tuning-forks and air-columns. Ten forks, giving a fundamental tone and its higher partial tones, are kept vibrating by electrical means, and with each fork a resonator, having the same period as the fork has, is associated. The sound of each fork is very feeble so long as its resonator is closed, but it at once becomes strong when the resonator is opened.

150. *Beats.—Dissonance.*—It has already been pointed out (§ 139) that the intensity of the resultant sound due to two component vibrations of slightly different periods varies alternately from a minimum to a maximum. These regularly-occurring maxima and minima constitute *beats*. One beat occurs in the time in which one vibration gains a complete period upon the other; and, consequently, the number of beats which occur per second is equal to the difference of the number of vibrations which take place per second in the two component tones. The following theory of the action of beats is that of Von Helmholtz.

When the beats succeed each other with too great rapidity, the ear becomes unable to distinguish them from one another, but it still recognises a distinct discontinuity in the sound, which produces a harsh effect known as *dissonance*. This dissonance is a maximum when there are about 38 beats per second.

The number of beats which occur per second when two given pure tones are sounded depends, of course, upon the absolute, as well as upon the relative, pitch of these tones. When the tones are in the neighbourhood of the middle C, maximum dissonance is produced when they differ in pitch by about half a tone; and there is almost no dissonance when the interval is a minor third.

It might seem, therefore, that dissonance should cease in all cases where the interval is sufficiently great. But, as Von Helmholtz showed, beating may occur between the partial-tones of two notes, and the result is that there may be considerable dissonance between the two, even although their fundamental tones differ by more than a minor third. Indeed, from this cause, none of the intervals below the octave form a perfect concord. Still, it is only in the cases of the second, and the major and minor sevenths, that the result is actually classed as a discord. In some cases the greatest dissonance, as above defined, does not produce the most disagreeable effect upon the ear.

Mayer has proved that the sensation of sound lasts for some time after the sound actually ceases, and that this period becomes smaller as the pitch of the sound increases. Hence, a sound may be intermitted a considerable number of times per second and yet appear to be continuous, the number of equal intermissions requisite for continuity being greater the higher the pitch is. Similarly, if beats occur sufficiently often per second the notes which give rise to these beats will appear to be continuous.

One of the most accurate methods of measuring the pitch of a note is based on the existence of beats. In Scheibler's *tonometer* a series of forks, of graduated pitches, is used, each fork giving four beats per second with its nearest neighbours in the series. By finding the number of beats per second which any note gives when sounded with each of the two forks which are nearest it in pitch—one being lower, the other higher—an accurate estimate of the pitch of the note may be made.

151. *Beat Tones.—Combination Tones.*—When the rapidity of the beats above described is sufficiently great, the effect upon the ear is that of a musical note having a vibration-frequency equal to the frequency of the beats. Tones due, in whatever manner, to beats, are termed *Beat Tones*. Von Helmholtz did not consider that there can be a true blending of beats so as to form a resultant

tone. As we shall see shortly, he assigned a different cause to a note apparently produced in this way. Earlier investigators believed that there is true blending of the beats, and their view is supported by Kœnig.

The tone which is produced under the conditions just considered is known as the *grave harmonic* of Tartini. It was originally discovered by Sorge in 1745, and was subsequently re-discovered by Tartini. Young called it a difference-tone, since its frequency is the difference of the frequencies of the original notes. To this tone Von Helmholtz added another, which he called a summation-tone, since its frequency was the sum of the frequencies of the original notes. Kœnig has shown that the latter does not exist when pure tones, such as those given by feebly-vibrating tuning-forks, are used.

According to Kœnig, there are two sets of beat tones, which are respectively due to two sets of primary beats—an *inferior* and a *superior* set. To find the frequency of the beats which will be produced by two pure tones of given frequencies, divide the higher of these frequencies by the lower, and take the smallest positive and negative remainders. The former remainder gives the frequency of the inferior set of beats, the latter gives the frequency of the superior set. For example, let the frequencies of the notes be 75 and 40 respectively. The quotient of 75 by 40 may be represented as 1, with a positive remainder of 35, or as 2, with a negative remainder of 5. In this case, therefore, the inferior beats have a frequency 35, while the superior beats have a frequency 5. On the other hand, notes of frequencies 75 and 80 give an inferior set of beats of frequency 5, and a superior set of frequency 70. The inferior set, when sufficiently rapid, give rise to Tartini's grave harmonic; and the superior set give rise to a superior beat tone. Kœnig has produced distinct beat tones from tones which themselves have too high frequency to be audible.

When a vibrating body (say the drum of the ear) is distorted to a slight extent merely during the vibrations, the reverse force, which tends to bring the body back to its position of equilibrium, is strictly proportional to the displacement. But this proportionality no longer holds if the displacement ceases to be small. Von Helmholtz showed that, in this case, if there are two independent causes of vibration, of frequencies n and n', in addition to vibrations of these frequencies, there will also be produced, among others, vibrations of frequencies $2n$, $2n'$, $3n$, $3n'$ and $n'-n$. The last of these frequencies corresponds to Tartini's grave harmonic, or Young's *difference tone.* In addition to these, a vibration of frequency $n'+n$ will occur, and this gives Helmholtz's *summation tone*. There will also occur difference tones and summation tones with reference to the original tones and the higher harmonics. To any such tone Von Helmholtz gives the general name *combination tone*, and he has proved that such tones occur when powerful primary vibrations are used.

The difference tone of frequency $n'-n$ has the same frequency as Kœnig's inferior beats; and the difference tone of frequency $2n-n'$ has the same frequency as Kœnig's superior beats. Yet, since Kœnig gets his beat tones under feeble excitation, the beat tones cannot be the combination tones. Voigt has given a theoretical deduction of Kœnig's law of beat tones.

Combination tones play an important part in Von Helmholtz's theory of dissonance, for beating may occur between a combination tone and a primary, or between two combination tones, just as it occurs between higher partial tones. For this reason also dissonance may occur in the case of two pure tones where the interval is greater than a minor third. Thus, when the interval is a major seventh, the higher tone will make 450 vibrations per second if the lower makes 240 per second. The difference tone has therefore 210 vibrations per second, which produces 30 beats per second with the lower primary.

Combination tones of a higher order may be produced between the first combination tone and either of the primary tones, and so on.

CHAPTER XIII.

LIGHT: INTENSITY, SPEED, THEORIES, COLOUR.

152. *Rectilinear Propagation. Intensity.* — In a homogeneous medium, light, in general, moves in straight lines. This may readily be seen if the light be passed through dusty air. The fact that the shadow of an object which is cast on the ground by sunlight is not defined with mathematical accuracy, but has a more or less blurred edge, does not disprove the statement. The indistinctness of the boundaries of the shadow is due to the finite size of the sun's disc, the light proceeding from each point of which produces a separate shadow. The more nearly the source approximates to a single point, the sharper does the boundary of the shadow become. This is well seen when the source is an electric arc.

The cases in which the above rule is departed from will be discussed in Chap. XVI.

The total intensity of a luminous source is measured by the amount of light which it emits per unit time, and the intensity of the light at any given point of the medium is measured by the quantity which falls per unit time upon unit area taken perpendicular to the direction in which the light moves at that point. And, since light moves out uniformly in all directions from a point-source in a homogeneous medium, it follows that its intensity varies inversely as the square of the distance from the source; for the same total quantity is, at different instants, spread over the surfaces of different concentric spheres, the areas of which vary directly as the squares of their radii. We assume, of course, that the light is not in any way absorbed in its passage through the medium. If any absorption did occur, the quantities of light passing through the different concentric spheres could not be equal.

In order that the above statement regarding the measurement of intensity may be of any practical value we must know exactly what is meant by light as a physical quantity. In popular language, we use the word vaguely with reference to the antecedent of a certain sensation; and we can get a rough idea of intensity by the extent to which the sensation is excited, though estimates made by different persons may differ greatly from each other. As a matter of science, we require to use measuring instruments whose indications are independent of the observer, and the nature of these instruments will depend upon the nature of the quantity to be measured.

We speak of light as being 'given out' by a source. It is evident, therefore, that an inquiry into the nature of its sources will be likely to lead to the discovery of something definite regarding its own nature. Such an inquiry being carried out, it is found that, in most cases, the source of light is a material body at a somewhat high temperature, the intensity of the light as judged by the eye being greater the higher the temperature of the body is. And, if energy be not supplied to the body from external sources, the intensity of the light will gradually diminish. This effect is, of course, in part due to the fact that the body is giving out heat as well as light. But if the rate at which it loses energy from this cause were measured, and if an excess of energy just sufficient to make up for this loss were supplied to the body, it would still be found that the intensity of the light diminished; and it would continue to diminish unless a sufficiently greater excess of energy were supplied to the body. We therefore are justified in making the supposition that light is a form of energy. And the truth of the supposition becomes confirmed when we find that the converse of the above phenomenon also happens—that light incident upon a body can produce increased motion of the particles of that body —and further that, in the radiometer (§ 129), visible motion of a considerable mass of matter may be caused by light.

The process by which light is given out is termed the process of *radiation*, and it is not unusual to employ this word as a name for the energy which is given out by that process. We shall see afterwards that the word, so defined, includes more than light, which might be called luminous radiation.

The strictly scientific method of measuring the intensity of light consists in measuring it as a quantity of energy transferred per unit of time across unit area. Thus we might transform it into heat, and measure it in terms of the quantity of heat required to raise the temperature of one gramme of water from 0°C to 1°C. This unit is called a *calorie*. The absolute unit is the *erg*—the rate per second at which work is done on a mass of one gramme moving with the constant acceleration of one centimetre per second. In such measurements it is usual to effect the transformation of light into heat by allowing it to fall on a surface coated with lamp-black (see § 234).

For practical purposes, an arbitrary standard is employed. In England this is a spermaceti candle of definite size and form. It is not a very trustworthy standard. The British Association recommends the adoption of a standard pentane lamp in preference. The Carcel lamp is used in France. The Violle standard unit, recommended by the French Congress of Electricians in 1884, is the light emitted per second from a square centimetre of the surface of melted platinum at the temperature of solidification. The Berlin standard, in which platinum foil at a definite temperature is used, is reproducible with an error of not more than one per cent.

Instruments used for the purpose of comparing the intensities of different sources of light are called *photometers*. In one of the simplest forms of a photometer, due to Count Rumford, the shadow of a rod is thrown upon a screen by each of two sources of light. A portion of the screen within any one of the two shadows receives no light from the source which casts that shadow, but is illuminated from the other source. Hence, if the two sources are equally intense, equally distant from the rod, and also equally distant from the screen, the shadows will appear equally dark. And if, under these conditions, the sources be not of equal intensity, the shadows are not equally dark, for then the quantity of light per unit area of the screen, which reaches one shadow from the one source, differs from that which illuminates the other shadow. If now we move the stronger source farther away from the rod we may make the shadows equally dark. And this occurs whenever, per unit area of the screen, equal quantities of light reach the two shadows. In this case the intensity of each source bears the same ratio to the square of the distance from that source to the shadow illuminated by it. It is desirable that the sizes of the shadows should be equal, and that the inclination to the screen of the lines joining the shadows and their respective sources should be similar. These conditions cannot be simultaneously attained when the sources differ in intensity.

In Bunsen's photometer a sheet of paper, upon which there is a greased spot, is used. Since light is transmitted readily through the greased portion, if the paper be illuminated from behind, the spot appears bright; if it be illuminated from in front, the spot seems dark; and, if it be illuminated to an equal extent from both sides, the spot vanishes, provided that the light is transmitted (or stopped) with equal readiness on both sides. Under the latter condition the intensities of the sources are directly proportional to the squares of their distances from the spot.

In practice it is not found that the spot vanishes under the same conditions when viewed from opposite sides of the screen. Observations are made on both sides, and a mean of the results is taken. In another form of the instrument two grease spots are employed, each being illuminated from behind by one source alone. The distances of the sources from the screen are varied until the two spots appear to be equally bright. In modern forms, a small part of the screen is made thinner than the rest, so as to be more transparent, and the grease spot is dispensed with.

Joly's photometer consists of two equal rectangular semi-transparent prisms of paraffin, separated by a thin opaque sheet. These prisms are placed between the two sources, which are to be compared, in such a way that each prism is illuminated only by the source which it faces. They are then moved along the line joining the sources until both seem to be equally illuminated.

The Lummer-Brodhun prism photometer is a nearly perfect realization of the principle of the Bunsen photometer. In the latter the grease spot should theoretically be perfectly transparent, while the rest of the screen should be perfectly reflective. In the former these conditions are practically attained by the use of two equal right-angled isoceles prisms of glass. The central portion of the hypotenuse face of one of these is left plane, while the remainder is slightly ground away, and the two are firmly bound together so as to form a compound prism of square section. If the eye be placed at one side of this prism, while a source of light is placed on the opposite side, no light will reach the eye from that source except through the central part, where the component prisms are in close contact, for total reflection (§ 163) takes place at the outer parts of the hypotenuse face of the prism next the source. And, if a second source be placed opposite the remaining short face of the prism which is next the eye, no light from this source will be reflected to the eye from the central parts of the hypotenuse face of this prism, while total reflection takes place from the outer parts. When the sources are at the proper distances from the centre of the square prism the whole field seems be equally illuminated.

The delicacy of the two latter photometers can be greatly increased by mounting the prisms on a spring. A slight vibration of gradually-decreasing extent greatly assists the eye in judging the position at which the adjustment is exact. In this way the Lummer-Brodhun photometer may be made to furnish values of relative intensity, which are correct to one-fifth per cent.

Various other forms of the instrument are in use. Varley suggested that the intensity of the stronger source may be cut down to that of the weaker by sufficiently rapid and extensive intermission. Leguy and Verschaffel used a radiometer (§ 129) in which the vanes have a bifilar suspension, so that a given deflection is produced by different sources when placed at different distances from the instrument.

All photometers such as those which have now been described give more or less accurate measurements of relative intensity, so long as the sources under comparison give out light of the same colour. But, when the colours differ, other methods must be used. O. N. Rood has pointed out that when the two halves of a rotating disc are illuminated to extents differing by not less than two per cent., a slight flickering is perceived when the rate of rotation has a suitable value. This flickering disappears when the intensities are equalized, although the colours may differ greatly. A. M. Mayer employs the ingenious method of making the different tints *appear* to be the same by the effect of contrast. Each of two thin discs is divided into sixteen equal sectors, and the surfaces are cut down to half their amount by the removal of alternate sectors. They are then strictly superposed, and a continuous disc of white translucent paper is clamped between them. The compound disc is then mounted vertically, with its plane bisecting the angle (150°) between two mirrors, and is set in rotation. By looking in the mirrors, the two sides of the disc can be seen simultaneously in close juxtaposition. One side is illuminated by one of the sources; the opposite side is illuminated by the other source. Suppose, for example, that daylight and lamp-light were being compared: the light passing through the translucent paper from the lamp makes the side illuminated by the daylight take a slightly yellowish tinge. Similarly, the side illuminated by the lamp has a slightly bluish tinge. A smaller disc, coloured in sectors in such a way as to appear orange yellow when in rapid rotation, is mounted on the side exposed to the daylight. This has, by contrast, the effect of diminishing the tinge of the surrounding ring of the original disc on that side to a very faint orange-yellow. A disc, so coloured as to appear greenish-blue when in rotation, is mounted on the other side, and this makes the surrounding (slightly blue) ring on that side appear, by contrast, to have the same faint orange-yellow tinge as the ring on the other side has. The apparatus is then used as an ordinary photometer. Further delicacy of indication is attained by making the central parts of the translucent sectors differ in translucency from the outer parts and by mounting a coloured ring, round the periphery of the compound disc, to act in the same way as the central coloured disc acts.

153. *Speed.*—In last section we have spoken of the *motion* of light. The use of the term is justified by the fact that a flash of light is not simultaneously seen by two observers who are situated at different distances from the source. When the distance between the two points of observation is not very large—a few miles, say—the interval of time which is occupied by the light in passing from one point to the other is so small that it cannot be measured except by very special means. Galileo first suggested the experimental investigation of the speed by means of a flash of light sent from one observer to another, stationed at a distance, who replied similarly at the instant when the light reached him. This experiment was carried out by the Florentine Academicians, the result proving that the speed was far too great to be detected by the method employed. Nearly two centuries then passed before the speed was experimentally determined by a direct process.

Before this was done, the speed had been determined by two independent astronomical methods.

The first of these is due to Roemer, who, in 1675, observed that the eclipses of one of Jupiter's satellites do not appear to recur at equal intervals of time, and pointed out that this would be a necessary consequence of the finite speed of light.

In order to understand more clearly how this may be, we may take an illustration from the phenomena of sound. If an observer be situated at a fixed distance from a point at which a gun is fired off at equal intervals of one minute, he will hear the report at equal intervals of one minute. But if, between two successive discharges of the gun, he move nearer to it, he will hear the next report at a shorter interval of time than one minute; while, if he move farther from it, the interval will necessarily be greater than one minute. (When applied to light (§ 178), this principle is usually called Döppler's principle.) The speed of sound may be determined from the results of two such observations. Let τ be the time which elapses between two successive discharges, and let t be the interval noted by the observer between two successive reports when he has meanwhile increased his distance from the gun by the amount d. Then t' being the (unknown) time taken by the sound to pass over the distance d, we get $t=\tau+t'$, and therefore the speed of sound is given by the quotient of d by $t-\tau$.

Now the eclipses of the satellite occur at instants which are very accurately calculable from known astronomical data. But the observed instants at which the eclipses apparently take place as seen from the earth do not always coincide with the calculated instants; and the errors at different times of the year are (assuming the finite speed of light) due to the variation of the distance between Jupiter and the earth. The greatest possible apparent error is the time which light takes to pass over the greatest difference of distance. But the greatest difference of distance between Jupiter and the earth is the diameter of the earth's orbit; and so we obtain the time taken by light to pass over this known distance.

Roemer's estimate of the time was 22 minutes. Later determinations showed that this value is too large and gave about 16 minutes 36 seconds as the true value. This makes the speed nearly 300,500 kilometres per second.

Roemer's conclusion regarding the finite speed of light was not accepted without question. D. Cassini and Miraldi raised objections, the former on the ground that the observed results might be due to a true inequality in the motion of the satellite, the latter on the ground that a similar inequality, depending on the position of Jupiter in his orbit, might exist and give rise to the phenomenon. These objections were, of course, quite legitimate.

The other astronomical method is due to Bradley. He, in 1728, observed that the fixed stars appear to describe small ellipses on the surface of the heavens in the course of a revolution of the earth round the sun, each star being displaced from the centre of its elliptic path *in the direction of the earth's motion in its orbit, and each to the same amount.* He concluded that this was due to the finiteness of the speed of light as compared with the speed of the earth in its orbit.

A simple illustration may make this clear. On a still day, raindrops fall vertically downwards. But, if one moves forward with considerable speed, they do not seem to fall vertically; they apparently fall in a slanting line, which is inclined forward from the vertical in the direction of the observer's motion. And it is evident that the apparent velocity of the drops is the resultant of their actual velocity and a velocity equal and opposite to that of the observer.

The light which comes from a star appears to come in a direction which depends in the same way upon the velocity of light and the velocity of the earth in its orbit. This latter velocity and the maximum angular displacement of a star from its true position being known, we can calculate the speed of light. The value which was found by Bradley for the angular displacement of a star from its mean position (called the constant of aberration) was $20''\cdot26$. This gives about 300,000 kilometres per second as the speed of light. Struve's determination of the value of the constant, made in 1845, gave 20″·445. Nyrén has recently given the value 20″·492 with a possible error of 6 in the third decimal place.

Struve's value led to an estimate of the speed of light which was smaller by one per cent. than the then accepted value of the speed as determined by Roemer's method. Now, it might be that this difference is due to the non-application of some correction to the assumption that the proper direction in which a telescope should be pointed in measuring the constant of aberration is the apparent direction in which the star lies because of aberration. If there is to be no bending of a ray of light when it enters a new medium which is at rest, the ray must pass along the normal to the surface of that medium. But the question still remains whether, when the new medium, say the object-glass of a telescope, is in motion, it is the direction of the ray as affected, or as not affected, by the motion which should lie along the normal. If it is the latter, a defect of one per cent. in the estimated speed might arise with ordinary telescopes which are pointed towards the apparent position of the star. Airy settled this point by an experiment, originally suggested by Boscovich, in which the tube of a telescope was filled with water. This made no difference in the observed value of the constant, and so proved that the thickness of the objective had no effect, and that the apparent direction of the star was the proper direction in which the telescope should be pointed. The phenomenon of aberration will be further considered in Chap. XXXII.

In 1849 Fizeau first determined the speed of light by direct experiment. Two telescopes were placed, at a distance of 8,633 metres apart, so as to directly face each other. In the eye-piece of the telescope at which the observations were made, a piece of transparent glass was fixed at an angle of 45° to the line of collimation, and a reflector was placed in the focus of the other telescope. A ray of light, reflected from the transparent glass, passed to the reflector of the second telescope, and was returned from it to the eye of the observer who looked through the first. A toothed wheel was placed near the eye-piece in such a way that the outward-going and the returning rays passed through a gap between two teeth. The wheel was then caused to rotate, and, at a certain rate of rotation, it was found that the light ceased to pass back between the teeth; the reason being that an adjacent tooth had moved into the place of the gap in the time that the light took to travel twice over the distance between the wheel and the mirror. The rate of rotation of the wheel, and the number of teeth which it contained, being known, the time which was taken by the light to pass over the given distance can be readily found. If N is the number of revolutions which the wheel made per unit of time, while n is the total number of gaps and teeth in its circumference, the speed of light is $2dNn$, d being the distance between the wheel and the mirror. When 2N revolutions are made per second the light is again bright, and so on. The value of the speed thus obtained was 314,000 kilometres per second.

Fizeau's experiments were repeated, in 1870 and 1874, by Cornu with improved apparatus; the result finally obtained by him was 300,400 kilometres per second. And, still more recently, a further improvement of the same method has been effected by Professor G. Forbes and Dr. J. Young.

Cornu's experiments were performed on a much larger scale than were Fizeau's, and he was consequently enabled to observe successive disappearances and reappearances of the light until fifteen teeth passed before the return of the light. These experiments made very evident the essential defect of the method; the extinctions and reappearances of the light are not sudden, but gradual, and so it is not very easy to judge when the extinction is complete or the light is brightest. The method of Forbes and Young was devised to obviate this difficulty to some extent. Another telescope tube with a reflector was placed at a greater distance than the former from the observing telescope so as to be almost in the same line. The first extinction in the far telescope occurred with a slower speed of the wheel than did the first extinction in the near telescope. Then, with increasing speed of the wheel, the intensity of the light in the far telescope increased while that in the near telescope decreased. Ultimately the two intensities became equal, and it was more easy to judge when this occurred than to judge when the extinction was complete or the light was brightest.

Wheatstone, soon after its invention, suggested that his rotating mirror might be used for the purposes of determining the absolute speed of light, and of comparing the speeds in air and in water. In 1838 Arago attempted to carry out the investigation. His plan was to send a flash of light simultaneously through air and water on to the revolving mirror, and to measure the relative deviations produced in the reflected rays. In this form it was impossible that the experiment could succeed, for the exact position of the mirror when the flash was emitted was not known, and the flashes seen through the water would reach the mirror at various stages of its revolution. Arago never overcame the difficulty. This was done by Foucault and Fizeau, almost simultaneously, in 1850. They allowed a beam of light from a slit to fall steadily upon the rotating mirror, and placed a fixed concave mirror in such a position as to exactly reverse the path of all rays reaching it from the revolving mirror. If the latter has turned through a small angle in the time taken by the light to pass twice over the distance between it and the other, the beam will not pass back through the slit, but will be deflected from it through a measurable distance. Thus Arago's difficulty was overcome by making the revolving mirror itself emit the flashes when in a definite position. And, if there were more than 30 revolutions per second, the separate flashes appeared to the eye to be continuous, so that the position of the deflected ray could be readily recorded. It was found that the speed was greater in air than in water. The absolute speed of light may be found in terms of the deflection of the ray and of the distance between the mirrors together with the rate of rotation of the revolving mirror. Foucault made the latter determination in 1862, after it had been seen that previous estimates of the distance of the earth from the sun were too great and that the true value might be found by an independent estimate of the speed of light. He measured the deviation of the image of a wire, not that of a slit. His result was 298,000 kilometres per second.

Wolf has recently suggested an improvement of Foucault's apparatus. The defect of the apparatus as used by Foucault was the smallness of the deviation produced with convenient speeds of the mirror. Bessel pointed out that the deviation may be increased by returning the deflected ray to the rotating mirror. Wolf proposes to realize this by placing the slit at the centre of the fixed mirror, giving both mirrors the same curvature, and placing the centre of the revolving mirror at the centre of curvature of the fixed mirror.

In 1878 and 1882 Michelson repeated Foucault's experiment in a form which allowed the deviation to be increased by increasing the distance between the mirrors. In Foucault's apparatus this distance was 4 metres; in Michelson's it was 605 metres. The fixed mirror was plane, and the reflected light was focussed upon it by a lens of great focal length. The other focus of the lens coincided with the image of the slit as seen in the revolving mirror. The result obtained in 1878 was 299,910 kilometres per second, that obtained in 1882 was 299,853 kilometres per second, as the speed of light in vacuo. The probable error was 5 kilometres per second.

In 1882, Newcomb made a similar experiment with the mirrors at a distance of 3,720 metres apart, the revolving mirror being a square prism of polished steel. He found the value 299,860 kilometres per second.

Michelson shows that the method of increasing the deviation by increasing the distance between the mirrors can be made much more accurate than the method of increasing it by multiplied reflections.

154. *Theories.*—We have already seen that transference of light implies transference of energy; and it is in this sense that we speak of light as a form of energy.

We are therefore limited to two suppositions regarding the physical nature of light. It may consist in the actual propagation of material particles, or corpuscles, from the luminous object; or, it may consist in the propagation of wave-motion through a material medium which fills space.

The former theory is known as the *Corpuscular Theory*, the latter as the *Wave Theory*, or *Undulatory Theory*, of light.

If the corpuscular theory were true, the mass of a corpuscle must be excessively small. For vision, according to this theory, is due to the impact of the corpuscles upon the retina; and the speed of these corpuscles is so great that, unless their individual mass were almost vanishingly small, the structure of the eye would be completely destroyed by the impact. The theory is met by a number of difficulties at the very outset. Thus it is somewhat difficult to account for the fact that the corpuscles have the same speed whatever be the temperature of the object from which they are projected. Again, the mass of a luminous body must be appreciably affected by the emission of particles; but there is no evidence of any such effect. Still, if we boldly overlook any such preliminary difficulties, we shall find that the theory enables us to account readily for many of the phenomena of light, although ultimately it fails us altogether.

On the wave-theory, vision is due to the communication of the vibrations of the assumed luminiferous medium (called the ether) to the nerve-ends of the retina. The molecules of a luminous body are (§ 176) in rapid vibratory motion, and this motion is communicated to the particles of the ether and is propagated through it from particle to particle, giving rise to a series of waves which travel with the speed of light. The investigation of § 136 has a direct application to the present case, and shows that the intensity of light is proportional to the square of the amplitude of vibration of the particles of the medium. The simplest motion of each particle is simple harmonic motion; the most complex is the resultant of several simple harmonic motions. Now, throughout a time which is long in comparison with the period, the energy of a particle vibrating in simple harmonic motion is half potential, half kinetic. To see this refer to Fig. 15. The total energy of the particle N, being constant, may be represented by OP^2. Its kinetic energy is on the same scale represented by NP^2. Hence its potential energy is represented by ON^2. But, in a complete period, ON goes through the same changes of length as NP. Hence the statement follows. And we can therefore assert that the energy contained in unit volume of the ether is, throughout a time which is long in comparison with the period, half potential, half kinetic. Further, if the wave-length of the disturbance be small in comparison with the unit of length, we may dispense with the restriction as to time, and assert that the condition holds at any instant. We shall find subsequently (Chap. XVII.) that the direction of vibration in the medium must be perpendicular to the direction of propagation of the waves. On the corpuscular theory the intensity must be proportional to the space-density of the corpuscles.

The wave-length is the distance, measured in the direction of propagation, from any point to the next point where the motion is similar. (Compare § 133.)

The wave-theory is not without its difficulties—many of them, indeed, are of a most formidable nature. But, as will appear, the evidence in favour of it is of such an overwhelming nature that its truth is now practically regarded as definitely established. Newton rejected it because he was unable to explain by it the rectilinear propagation of light. It is now known that the existence of rays is a necessary consequence of the fundamental principles of the theory. The special form of the wave theory known as the Electromagnetic Theory will be discussed in Chap. XXX.

155. *Colour.*—Many different kinds of light are recognised. We speak of red light, blue light, etc. On the corpuscular theory the difference must be inherent in the corpuscles. On the wave-theory the difference is a mere difference of wave-length, or of vibrational period—which is only another way of stating the same thing, since the speed of propagation of light of all colours has, in free space, one definite value only.

In addition to their difference in respect of tint, two colours may differ in respect of luminosity and of purity. The latter quality refers to admixture with white light, the purity being less as the amount of white light present is greater. Sunlight or ordinary daylight is the standard of white light. The colours present in sunlight (or in any light) may be exhibited by means of a prism in the manner described in § 171. The diagram of § 178 gives the scale of the visible spectrum. The line A shown in that diagram is in the dark-red part of the spectrum, B and C are in the brighter red part, D is in the orange-yellow part, E and *b* are in the green part, F is near the borders of the green and blue parts, G is near the borders of the blue and violet parts, and H is in the deep violet part. It is customary to speak of the six colours just mentioned as the constituent colours of sunlight, though the ordinary eye can distinguish about 300 colours. This has been proved by Dr. Kœnig. He intercepted, by means of a diaphragm with a narrow slit, all the light from one spectrum, except that in a definite, very narrow region. The same thing was done with another equal slit and a similar spectrum. This slit was then moved along the second spectrum until the eye could just distinguish between the hues of the light passing through the two slits. This determined the sensitiveness of the eye to the colour in the given region of the spectrum. Theoretically, of course, it is only light of one definite wave-length which is monochromatic, *i.e.*, of one hue.

When the intensity of coloured light is sufficiently great, the eye loses all power of distinguishing hue. This power is also lost when the intensity is sufficiently feeble. This explains the fact that a hot body appears of a steel-gray colour before it becomes hot enough to appear dull red. By throwing patches of white light and of practically monochromatic light from a definite region of the spectrum side by side upon a screen in a dark box, and then diminishing the intensity of both until each seemed to have the same hue, Abney has estimated the intensity to which various parts of the spectrum must be reduced before they will appear to be colourless. The colourless light at the extreme (normally) red end of the spectrum is considerably more intense than that at the central regions, but by far the greatest intensity of colourless light is found at the (normally) violet end.

Abney has also determined the reduction of intensity which various spectrum colours, made originally of equal intensity, must undergo before the eye ceases to perceive any light. The whole of the blue and violet part of the spectrum requires about equal reduction, and that to about 25 times the reduction required by the green part, which has itself to be reduced about 11 times more than the red end.

With regard to purity, Abney finds that the eye cannot distinguish hue in any case if the colour is mixed with white light 75 times more intense than itself. In many cases less impurity than this will destroy the colour. The addition of a very small amount of white light to blue will be perceived, while a large relative amount may be added to yellow without apparently making any change.

Colour-mixing.—Various colours may be added together so as to produce a resultant effect upon the eye. This mixing of colours must be carefully distinguished from the mixing of pigments (see § 179). A simple method of mixing ordinary colours is to rotate rapidly a circular disc which is divided into sectors coloured in some given manner. The eye cannot separate the colours when the rotation is sufficiently rapid. Another method consists in placing sectors of coloured glasses in front of a lens which is used for projection on a screen; for each portion of the lens produces an image which is identical with that produced by the whole lens, except that the luminosity is less in the ratio of the area of the given sector to the area of the whole lens.

But, if practically pure spectrum colours are to be mixed, a more perfect arrangement must be adopted. Abney uses a modification of Maxwell's method. Instead of throwing a spectrum from a prism directly upon a screen as Maxwell did, Abney throws it upon a convex lens, which forms an image of the face of the prism upon a screen. If all the spectrum falls upon the lens, the colours are re-combined and the image is white. If a screen with a slit is interposed between the prism and the lens so as to allow one definite colour alone to reach the lens, the image will appear of that colour; and, by using a number of such slits, various colours may be combined.

Vision.—Nerve-fibres spread out from the optic nerve over the inner surface of the retina and communicate with several layers of ganglionic cells placed outside them. The perceptive layer is placed farthest out. The outer ends of the cells in this layer are either rod-shaped or cone-shaped. The layer is therefore often called the rod-and-cone layer. The base of a cone and one end of a rod are directed inwards. The inner portion of both is transparent in the human eye. The outer portion of a rod projects beyond the outer portion of a cone and contains a purplish-red colouring matter, called 'visual purple,' which is readily bleached under the action of light. Light which enters the eye passes through all these layers in order to reach the extremities of the rods and cones, which are the portions effective in vision. Beyond the rods and cones there is a region occupied by dark pigment cells which absorb light that has passed through the other layers.

The cones are more sensitive visually than are the rods. Now, there is a central region of the retina, known as the *yellow spot*, where no rods occur, where the cones are closely packed, and the other inner retinal layers are absent, while the nerve-fibres do not cover it, but pass round it. This is, therefore, the part of the retina which produces most perfect colour vision, and we employ it when we look directly at an object. The rods do not enable us to distinguish colour. There is a region surrounding the yellow spot where rods and cones occur with about equal frequency. In the more remote regions of the retina, the cones occur more and more seldom; and they are entirely absent from the extreme region. Hence, the power of perceiving colour diminishes from the central region outwards and vanishes in the extreme region. The part of the retina where the optic nerve enters the eye is devoid of both rods and cones, and is known as the 'blind spot.'

Langley estimates that the sensation of crimson can be produced by the expenditure of energy to the extent of one-thousandth part of an erg, while the sensation of green can be produced by the expenditure of only one-hundred-millionth part of an erg.

Stimulation is not instantaneous. A very faint light may require to act for half a second in order to produce its full effect. One-tenth of a second is about sufficient for light of ordinary intensity. And vision persists after the light ceases. This subject was partly investigated by Plateau. E. L. Nichols found that yellow light has less effect in this way than light from the ends of the spectrum. He also found that the effect of a given colour is less when the intensity and the time of exposure are large.

Two objects may subtend so small an angle at the eye that they appear as a single object. In this case it is found that, if the light from one object falls upon one cone, the light from the other object cannot fall upon an adjacent cone. For this reason the central part of the retina gives most acute vision, though it is not the most sensitive region as regards intensity of light.

Colour-blindness.—If we place three small brightly-coloured discs—one red, one green, and one blue—near each other, and look directly at the group with one eye, the colour of each can be distinctly seen. If we remove the discs sufficiently far from the direct line of vision for the light from them to pass off the yellow spot, the colour of the green disc ceases to be apparent; and, as we have already found, no colour can be distinguished when the light falls near the borders of the retina. The region outside the yellow spot is therefore blind to green, while the extreme region cannot appreciate any of these colours. Even the central parts of the retina in some eyes show these peculiarities. Some eyes can only appreciate variations of intensity, and others are blind to violet light. Most ordinarily, an inability to distinguish red and green is apparent. Occasionally, one eye is normal, while the other is colour-blind, and this indicates that the defect really lies in the eye and not in the brain; for a unilateral brain defect would influence corresponding halves of both eyes. The latter type of blindness actually results from brain injuries, all colour-sense being lost in the regions which are affected.

The spectrum seen by a colour-blind eye necessarily exhibits considerable differences from the normal spectrum. A narrow colourless band is always present in it. Dr. Kœnig finds that the position of this band can be very accurately determined by means of the apparatus, described above, which he employs in the estimation of colour-sensitiveness in normal eyes. As a means of detecting colour-blindness, he employs the following arrangement: Two rectangularly polarized beams of light (Chap. XVII.) are passed through a quartz plate, and are then viewed through a Nicol's prism. When viewed by a normal eye, the two beams are always differently coloured, and the colours change if the prism is turned round its axis. When viewed by a colour-blind eye, there is always some position of the prism such that the two beams have no apparent difference of colour. The position of the prism at which this occurs depends on the thickness of the quartz plate, on the intensity of the light, and on the peculiarities of the eye.

The question of colour-blindness will be further considered when we deal with the theories of colour-vision.

Complementary Colours.—Newton found that various pairs of spectrum colours may, with normal vision, be combined so as to form colourless light. Thus light from a red part of the spectrum, when mixed with light from a greenish-blue part, gives white or grayish light as the result. Such pairs of colours are said to be complementary. The colour-blind outer portions of the retina cannot, of course, discover these pairs. The parts of the retina nearer the yellow spot, and eyes similarly blind to one colour, give complementary pairs which may be formed from parts of the spectrum taken closer and closer together and finally coinciding at the neutral point. In normal eyes the approach of the colours cannot be carried to a point of coincidence.

Contrast Colours.—If in a partially-darkened room a beam of coloured light be caused to throw a shadow on a white screen, a normal eye will perceive the shadow to be illuminated by the complementary colour. Thus, red light will give a greenish-blue shadow, and a purple shadow will be produced by greenish-yellow light. The colour of the shadow is said to be produced subjectively by contrast with the stronger light surrounding it. Another example has already been given in the discussion of colour-photometry. Stilling states that a contrast colour cannot appear if the eye is unable to distinguish between the colour of the light that is used and the complementary colour. Hence, this method may be used to detect inability to distinguish red and green.

Young-Helmholtz Theory.—Just as white light may be produced by the combination of all the spectrum colours, so various spectrum colours may be produced by the combination of others. Young remarked that 'it is certain that the perfect sensations of yellow and of blue are produced, respectively, by mixtures of red and green and of green and violet light, and there is reason to suspect that those sensations are always compounded of the separate sensations combined; at least, this supposition simplifies the theory of colours; it may, therefore, be adopted with advantage, until it be found inconsistent with any of the phenomena; and we may consider white light as composed of a mixture of red, green, and violet only, in the proportion of about two parts red, four green, and one violet, with respect to the quantity or intensity of the sensations produced.' As early as 1722, Mayer had referred all colours to red, yellow, and blue.

Young assumed that the retina contained three sets of fibres, each set being capable of responding to one of the three fundamental colours alone; and that all other colour sensations were due to the stimulation of one or more of the three fundamental sensations in various degrees. Some such hypothesis as Young's is necessary in order to avoid the difficulty of assuming that there is in every small part of the retina a separate mechanism consisting of parts, each of which is capable of being affected by one definite luminous undulation alone. Von Helmholtz suggested that the retinal terminations of the three sets of fibres might contain a different substance capable of decomposition under the action of light; and that, for example, under the action of red light, one of these substances is more extensively decomposed than are the others, thus giving rise to relatively great action in the corresponding set of nerve fibres.

The above three standard colours are chosen because no one of them can be matched by mixtures of colours from other parts of the spectrum. In his experiments on colour-mixing, Maxwell chose as the standard red a bright scarlet at about one-third of the distance from the line C to the line D in the spectrum; as the standard green, he chose light near the line E; and, as the standard violet, he chose light at about one-third of the distance from F to G. And these colours are not strictly Young's standards.

Maxwell represented all colours in terms of these three standards, after Newton's method, by supposing equal quantities of the standards to be placed respectively at the corners of an equilateral triangle; and, in the case of any other colour composed of these in given proportions, finding the centre of inertia of masses having these proportions and supposed to be placed at the corners. This point represented the given colour. Otherwise, the nature of a compound colour may be represented on the colour pyramid, after Lambert's method, by measuring, along three lines at right angles to each other, lengths proportional to the quantities of the three standards which are present in the given compound colour. It is interesting, in this connection, to note that Abney and Festing have shown that any colour can be matched by diluting some pure spectrum colour with white light, so that any colour, howsoever complex, can be represented in terms of wave-length and impurity.

Any linear relation, representing a given colour as a sum of definite quantities of the standards, constitutes a 'colour equation.' By forming colour equations for all the constituents of the spectrum it is possible to find how the three fundamental colour-sensations are called into action at various parts of the spectrum. This has been done by Dr. Kœnig. He finds that, if the sensations be represented by ordinates marked off perpendicular to the length of the spectrum, the curve of the red sensation starts from zero at the red end of the spectrum, rises to a maximum at a point rather nearer to the line D than the line E, and vanishes at a point about one-third along from G to H; the curve of the green sensation begins about the line C, reaches a maximum rather greater than the red maximum at a point nearer E than D, and vanishes at a point nearer G than F; and the curve of the violet sensation begins about D, rises slowly at first, and then, rapidly reaching a maximum higher than the others at a point midway between F and G, falls to zero at the extreme violet end of the spectrum. The red and green curves intersect at a point, beyond D, fully one-quarter of the distance along to E; the red and violet curves intersect at a point nearly half-way along from *b* to F; the green and violet curves intersect at a point about two-thirds of the way from *b* to F.

The theory gives a ready explanation of the fact that all sufficiently intense colours cannot be readily distinguished from each other or from white. In those parts of the spectrum which stimulate all the sensations, each sensation is stimulated then to its full extent. And as Von Helmholtz pointed out, the law of superposability ceases to be true when the luminosity is excessive. The uncertainty of tint is a phenomenon of the same kind as the uncertainty of absolute luminosity. And, with regard to weak light, Abney finds that the curve of luminosity, when the spectrum exhibits a gray appearance only, is identical in form with the curve of the violet sensation.

Colour-blindness, on this theory, is due to the absence of one or more colour sensations—red blindness to the absence of the red sensation, and so on. If the red sensation is absent from a certain eye, there should be a neutral point in the spectrum at the place where the violet and green sensation curves for that eye intersect. This is at least very nearly the case, though, on the other hand, Dr. Kœnig asserts that the neutral point for both red and green blindness is the same. Again, in both red and green blind eyes, the steel-gray spectrum of light of feeble intensity should be identical with that for normal eyes; and Abney finds that it is so. He also finds that the difference between the luminosity curve of a violet blind eye and that of a normal eye gives the curve of the violet sensation, and that the luminosity curve of a red-green blind eye agrees in general form with the curve of the violet sensation.

Contrast colour is explained in this way. Take, for example, the case of a candle throwing a blue shadow in daylight upon a white screen. The shadow is really white, while the rest of the screen is yellowish; but, relatively to the rest of the screen, the shadow contains more blue, for it has less yellow; and the eye takes this into account, referring them to a mean between the two as the standard of whiteness for the screen.

Hering's Theory.—Hering suggests that all visual perceptions are due to six fundamental sensations grouped in three pairs—white-black, red-green, yellow-blue. The sensation of white follows the decomposition of a certain substance in the visual cells, the sensation of black follows its restitution; red and green lights cause decomposition or recombination of another substance which must exist normally in visual cells in a state of partial decomposition; blue and yellow act similarly. The cells concerned in the white-black impressions are supposed to exist to a greater or less extent in all parts of the retina.

According to this theory, the sensation of white light may exist without the presence of any colour sensation, and it is stimulated by light of all kinds, while, on the other hand, no colour sensation can exist without the presence of some white sensation. Now Hillebrand, using an apparatus in which one half of the field of vision could be filled with white light while the other half was filled with monochromatic spectrum light of moderate intensity from the same source, found that, if he carefully rested one eye for some time before looking at the monochromatic field with it, the monochromatic field could be made to appear colourless, and could be made to match the white field by suitable change of intensity of the latter. In this way he could find the luminosity of the spectrum relatively to white light of given intensity, and the maximum was in the green—not in the yellow, as it is when tested by the unrested eye. If the matched fields were viewed with the unrested eye, a difference of luminosity became at once apparent. In the red and yellow parts of the spectrum, the coloured half of the field was more luminous than the white half; in the green and blue parts, the coloured half was less luminous than the white half. On Hering's theory this is explained by the statement that each part of the spectrum produces the impression of luminosity due to white light, together with luminosity inherent in the special colour sensation involved, which is positive in the red and yellow parts of the spectrum and negative in the green and blue. Further, Kœnig found that, in the case of an eye blind to red, yellow, green, and blue colours, the maximum luminosity lay in the green part of the spectrum, at the line *b*, where the maximum of Hillebrand's curve lies, and not where the maximum of the Young-Helmholtz violet sensation lies; and Hering also found that the luminosity curve of a totally colour-blind eye was identical with that of the normal eye when rested as described.

On this theory, red blindness is not recognised as distinct from green blindness: both arise from absence of the red-green sensation, and the apparent distinction between red and green blindness is due to the fact—which is evident in normal eyes, and is due to absorptive action in the eye, especially in the yellow spot—that some eyes respond more readily to yellow stimulus than to blue, while others have the blue sense more powerful than the yellow. Thus, the red of the so-called green blind eye is really yellow, and becomes neutral when diluted with blue; and the green of the red blind eye is really a blue, which becomes neutral when diluted with yellow. But Von Helmholtz did not consider that the absorptive action in the eye is sufficiently powerful to account for the observed differences.

Contrast colours are explained by the assumption that stimulation of one part of the retina by light of one colour causes in neighbouring parts, by some kind of inductive action, stimulation which gives rise to the sensation of the complementary colour. A. M. Mayer has shown that, when the light of a candle is made to throw an apparently blue shadow on a white surface in daylight, an observer, if placed behind a screen and allowed to look at the surface through a tube placed in the screen which enables him to see a portion of the shadow and a portion of the rest of the surface as two spots in close juxtaposition, will at once say that the portion of the shadow is blue and that the other portion of the surface is yellow, although he has been entirely misled as to what to expect, and does not know that the screen is white. He considers this to be an indication of the truth of Hering's theory.

As Von Helmholtz pointed out, Hering's theory is really Young's theory with a different choice of the three fundamental sensations· Hering's red and green are complementary, and therefore differ from Young's red and green, which are not complementary. Von Helmholtz showed that Hering's standards are connected linearly with Young's by the equations $u\sqrt{3}=x+y+z$, $v\sqrt{2}=x-z$, $w\sqrt{6}=x-2y+z$: where x, y, z, represent respectively Young's red, green, and violet; u represents Hering's white, $+v$ and $-v$ represent respectively his yellow and blue, and $+w$, $-w$, represent respectively his red and green. Thus, the positive sensation u, and the positive or negative sensations v and w, involve essentially the three necessarily positive sensations x, y, and z; and the two theories are physically identical, the only question at issue between them being whether it is the one set of sensations, or the other set, which have a physiological basis in the structure of the eye. And, in testing the point, it is essential to make experiments, under similar conditions, upon a large number of eyes. Kœnig found that the position of the neutral point in the spectrum of a colour-blind eye depends upon the intensity of the light; and this means that the sensation curves are variable to some extent in any eye when the intensity varies. Hillebrand's colourless luminosity curve was made with higher intensity than was Abney's curve of colourless luminosity. In all cases, the psychical aspect of the question has to be considered, and the effect of training also. Thus, as Helmholtz remarked, though the centre part of the retina is trichromatic while another part is bichromatic and the remainder is monochromatic, a white surface is always judged to be white on whatever part of the retina its image falls. The eye has learned by experience always to judge a surface as white if it would appear white when its image fell on the central part of the retina.

Von Helmholtz stated also that, while Hering's pure colour sensations can never be realized, since the white sensation is necessarily stimulated simultaneously, Young's pure colour sensations can be at least closely approximated to by the method of retinal fatigue.

Young's hypothesis is, of course, the simpler of the two.

Lodge's realization of an electric analogue of visual action will be considered in Chap. XXX.

CHAPTER XIV.

LIGHT: REFLECTION, REFRACTION, DISPERSION.

156. *Laws of Reflection.*—When a ray of light reaches the bounding surface of a homogeneous medium through which it is passing, it is, in part at least, bent back or reflected, and pursues a different, though still rectilinear path.

The reflected and incident rays lie in one plane with, and make equal angles with, the normal to the surface.

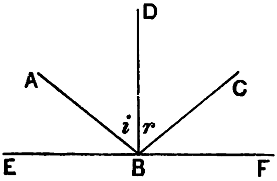


FIG. 64.

Let EBF (Fig. 64) represent a section of the bounding surface by the plane of the paper, and let BD be the normal to the surface at the point B whereon the incident ray AB falls. Then, BC being the reflected ray, the angles i and r, which AB and BC make with BD, are equal, and AB, BC, and BD all lie in one plane, which is normal to the reflecting surface. The angles i and r are called, respectively, the angle of incidence and the angle of reflection.

The law may be readily proved by observations made on a beam of light reflected from a plane mirror which can be rotated about a vertical axis while the incident beam preserves a fixed horizontal direction.

Many surfaces, such as those of chalk or of rough white paper, *scatter* or *diffuse* the incident light in all directions. But this is merely a special case of reflection. At every point of such a surface rays are reflected in accordance with the above law; but the whole surface is practically made up of an excessively great number of very small planes, which are indiscriminately inclined in all possible ways.

The intensity of the reflected ray depends upon the angle of incidence, being greatest when the angle is large, and having its least value when the incidence is perpendicular. With a given intensity of the incident ray it varies with the nature and the state of surface-polish of the reflecting substance, and it depends also upon the nature of the medium through which the light is travelling.

At perpendicular incidence, mercury reflects 75 per cent., water reflects 2 per cent., flint-glass reflects nearly 7 per cent., and plate-glass reflects 4 per cent., of the incident light. In the last case the intensity of the reflected light may exceed 80 per cent. of that of the incident light when the angle of incidence approaches 90°.

157. *Reflection from Plane Surfaces.*—If a ray of light be

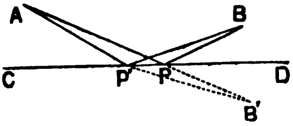


FIG. 65.

emitted from a point B (Fig. 65) and reach a point A, after reflection from a plane surface, CD, the actual length of the path APB is the shortest possible consistent with the condition of reflection at the given surface.

For an eye placed at the point A will see the light in the direction AP as if it came from a point B′, which is situated on the normal drawn from B to the surface.

This is so since the eye sees an object by means of a cone of rays: and the angle of the cone is unaltered by reflection since (Fig. 66)

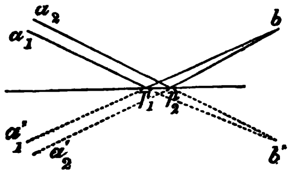


FIG. 66.

the angles which a_1p_1 and a_2p_2 make with the reflecting plane are respectively equal to the angles which $p_1a'_1$ and $p_2a'_2$, the continuations of the lines bp_1 and bp_2, make with that plane. And so the continuations of a_1p_1 and a_2p_2 meet at a point b', which is such that bp_1b' and bp_2b' are both bisected by the surface; and therefore b' and b lie on the same normal to the surface, and are equally distant from it.

And any other path, AP′B, being equal to AP′B′, is greater than APB, which is equal to APB′.

The point B′ is called the *image* of the point B. If B were a body of finite size, each point of it would give rise to an image; and the whole congeries of these point-images constitutes the image of the body B.

If the mirror can revolve about a fixed axis through P while B is fixed, the image B′ will always lie on a circle with centre P and radius PB. For CD always bisects BB′ perpendicularly, and this is a property of the normal from the centre to the chord of a circle. The reflected ray turns at twice the rate at which the mirror turns. The latter principle is employed, as in the sextant, for the measurement of the angular distance between two objects.

If two parallel mirrors face each other, and an object is placed between them, an infinite series of images will be formed in each. One image is formed directly in each mirror; each of these images gives rise to a secondary image, and so on. The position of the successive images can be readily ascertained by an application of the above law. Each image is dimmer than the preceding in geometrical progression, because of the loss of light on reflection. If the two mirrors are inclined at an angle to each other, the successive images must, for the reason given above in the case of the revolving mirror, lie on a circle whose centre is at the foot of the perpendicular drawn from the object to the line of intersection of the mirrors and whose radius is the length of that perpendicular. If the angle between the mirrors is an aliquot part of a right angle, there can only be a finite number of separate images, for th

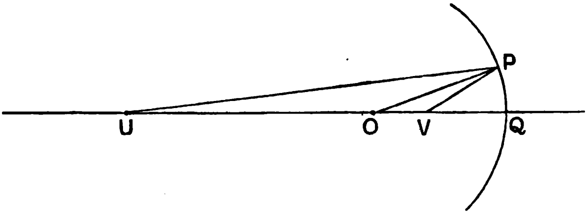


FIG. 67.

members of various sets are superposed. This principle is employed in Brewster's *kaleidoscope.*

158. *Reflection from Curved Surfaces.*—Let Q (Fig. 67) be the section of a spherical mirror by the plane of the paper. Let O be the centre of the sphere, and let a ray, UP, emitted by a luminous object at the point U, be reflected to the point V. We have $PVQ = UPV + PUQ = 2UPO + PUQ$. Therefore $PVQ + PUQ = 2(UPO + PUQ) = 2POQ$. When P and Q are nearly coincident this becomes approximately $PQ/PV + PQ/PU = 2PQ/PO$ or $1/PV + 1/PU = 2/PO$. If we denote by u, v, and r, the lengths of the lines PU, PV, and PO respectively, this gives

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}.$$

This equation enables us to calculate the position of the image V when the position of the object U is given. If U be situated at infinity towards the left-hand side of the diagram, V is half-way between O and Q. This point is called the *principal focus* of the mirror. As U moves in from infinity, V moves out to meet it, and the two points coincide at O, the centre of the sphere. The positions of U and V are now interchanged, and finally, when U is at the principal focus, V is at infinity towards the left. Whenever U comes nearer Q than the distance of a semi-radius, the quantity v becomes negative, that is, the image passes away (Fig. 68) behind the mirror, and gradually approaches it from infinity in this direction until both object and image coincide at Q.

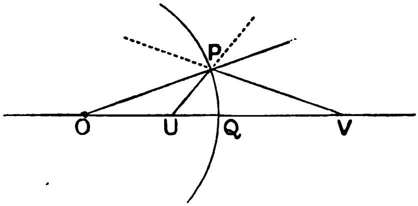


FIG. 68.

In all cases V and U are interchangeable, that is to say, V may be the object and U will then be the image. A slight inspection of the two diagrams will make this clear. Such points are termed *conjugate foci.*

The image, when on that side of the mirror on which the object is, is called a *real* image; when on the opposite side, it is called a *virtual* image.

It is evident that an object on the convex side of the sphere can have a virtual image only. In this case the above formula becomes

$$\frac{1}{v}-\frac{1}{u}=\frac{2}{r},$$

which is the modification of the formula necessary to make it apply to the case of reflection from a convex spherical mirror, since u

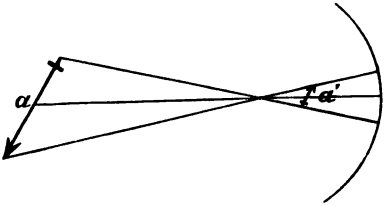


FIG. 69.

must be regarded as a negative quantity if v be taken as positive —the two quantities being measured in opposite directions from the mirror.

The case of a parabolic mirror is of special importance. If a luminous point be placed at its geometrical focus, the rays, after reflection, will proceed in lines parallel to the axis, for the normal drawn to the surface of the mirror at any point makes equal angles with lines drawn from that point to the focus, and parallel to the axis, respectively.

The law—stated in last section with reference to a plane mirror—that light takes the shortest possible path between two points con-

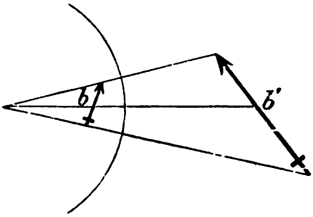


FIG. 70.

sistent with the condition of reflection at the given surface, still holds in the case of *any* surface provided that we limit the statement to other paths which do not finitely differ from the actual path of the light. The necessity for this limitation will be evident if we consider that light diverging from a focus of a reflecting ellipsoid may take the longest possible, as well as the shortest possible, path.

Figs. 69 and 70 show positions of real and virtual images: a and a' are mutually real images; b and b' are mutually virtual images; b being the image of b' in a convex mirror, while a, a', and b' are the images of a', a, and b in a concave one. When formed by one reflection, or by an odd number of reflections, a real image is inverted, but a virtual image is not inverted. The positions of the various points of the image corresponding to given points of the object are found by means of the above formulæ.

159. *Reflection Caustics: Focal Lines.*—Let CBQ (Fig. 71) represent the section of a spherical mirror by the plane of the paper,

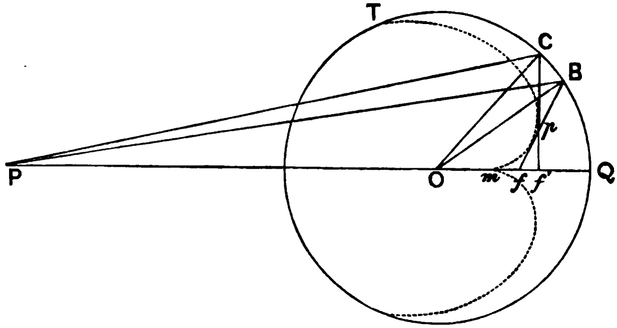


FIG. 71.

and let PC, PB represent two rays which, diverging from the point P, fall upon the mirror: let also the reflected rays, Cpf' and Bpf, intersect in the point p.

For *every* position of the point B in the circle we have BOQ= OPB+PBO=OPB+OBf. Hence, for two successive positions, B and C, the change in BOQ is equal to the change in OPB together with the change in OBf. Similarly, the change in BOQ is equal to the change in BfQ *less* the change in OBf. Hence twice the change in BOQ is equal to the change in OPB together with the change in BfQ: that is 2COB=CPB+CpB.

The results of last section follow as a particular case of this.

Let CB be an infinitesimally small arc of constant length, and let T be the point at which the tangent from P meets the circle CBQ. CPB always diminishes, and therefore (by the above equation) CpB constantly increases as C moves from Q towards T. Hence the length of Cp always diminishes as its inclination to CO increases, until finally p coincides with T.

The locus of p is called the *caustic curve*, and is indicated by the dotted curve in the figure. It touches the circle at the point T and the line PQ at a point m, the position of which may be found by the formula of last section.

If the whole figure be rotated about the line PQ, the circle CBQ traces out the spherical reflecting surface, and the caustic curve traces a continuous surface which is called the caustic surface. All points on this surface are more intensely illuminated by the reflected light than any point which does not lie upon it, and the cusp (at m) is the place of most intense illumination. All rays which are reflected from the surface pass through the line PQ.

The section of such a caustic surface is well seen on the surface of milk contained in a polished bowl.

Let us suppose that a small, but finite, circular cone of rays falls upon the reflecting surface in the neighbourhood of the points B, C. All rays from points on the small circle, the pole of which is Q, and which passes through B, intersect PQ on the point f; and all rays from points on the similar circle through C pass through f'; and so on. It is evident, therefore, that a plane which is perpendicular to the axial line of the reflected cone, and which passes through the point in which the axial line intersects the line PQ, will cut the cone in an elongated figure-of-eight-shaped area, which may be regarded as a straight line, and is called the *secondary focal line.*

Again, a plane drawn perpendicular to the axial line through the point in which that line touches the caustic surface, cuts this surface in a circle, and all the reflected rays will pass through the plane in the immediate neighbourhood of a small, practically straight, portion of the circle; so that there is another, nearly linear, normal section of the reflected cone. This is called the *primary focal line.*

The two focal lines are mutually perpendicular.

An approximately circular section exists between the two linear sections. This is called the *circle of least confusion*, and is the place where the reflected light most nearly converges to a point.

The defect, which spherical mirrors have, of not bringing all rays diverging from one point to a definite focus at another point, and of bringing some rays from different points to a focus at one point, is called *spherical aberration.* The distance, measured along the axis between the principal focal lines, is called the *longitudinal* spherical aberration, and the radius of the circle of least confusion is called the *lateral* spherical aberration.

160. The law of reflection follows readily from the principles of

the corpuscular theory. Let *pq* (Fig. 72) represent the path of a corpuscle, and let AB be the surface from which the corpuscle is reflected. When the corpuscle comes within a certain small distance from the surface, indicated by the line *ab*, it experiences the attraction of the medium which is bounded by the surface AB, and so is bent from its rectilinear path. The mutual action of the particle and the medium may alternate from attraction to repulsion many times according to an unknown law, but it must ultimately be a repulsion which (at *r*) stops the motion of the corpuscle towards the surface. The mutual forces still acting as before, the particle must now describe a path *rq'* precisely similar to *rq*, until, at *q'*, being freed from the action of the reflecting medium, it describes a rectilinear path, *q'p'*, which is inclined to AB at the same angle as *pq* is. And, since the action of the medium is everywhere in lines perpendicular to the surface AB, the particle retains its velocity parallel to AB unaltered during the process of reflection, and the lines *pq* and *p'q'* are in one plane with the normal to the surface.

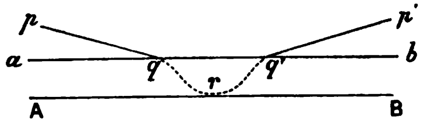


Fig. 72.

161. The wave-theory also affords a ready explanation of the phenomena of reflection.

But, before dealing with this point, it is necessary to consider the explanation, on this theory, of the rectilinear propagation of light.

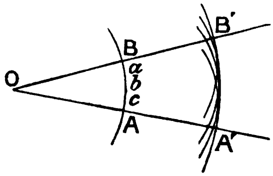


Fig. 73.

Newton did not see how to account for it, and so supported the corpuscular theory. Huyghens was the first to show that the wave-theory furnishes a ready explanation of the phenomenon.

Let AB (Fig. 73) represent a portion of a spherical wave-front diverging from the point O. All points, such as *a*, *b*, *c*, on this surface become centres of disturbance from which secondary spherical waves diverge. With radius AA′, or BB′, equal to the distance which light will travel in a certain time, t, describe circles from a, b, c as centres. These circles will all touch another spherical surface, A′B′, concentric with AB. This constitutes the new wave-front. At all points of this surface secondary wavelets are superposed, and so a strong resultant effect may be produced. At no other points, besides those on A′B′, are the effects of the separate wavelets superposed, and the isolated wavelets are too feeble to produce light. Hence the rays included in the region AOB diverge outwards in straight lines.

The above explanation is due to Huyghens, and will suffice for our present purpose. But, as Fresnel pointed out, Young's principle of interference is essential, in addition to the above, in order to make the demonstration rigorous. See § 200.

We shall assume for the sake of simplicity that we are dealing with a plane wave propagated through the luminiferous medium. Let ADF (Fig. 74) represent the reflecting surface, and let the given disturbance have reached the position ABC; so that ABC represents a portion of the plane wave-front, to which the rays (of which three are indicated in the figure) are everywhere perpendicular, the medium being assumed to be homogeneous and isotropic.

When the wave reaches the point A, the particles of the ether at that point are set in vibration and give rise to a spherical wave which spreads out from A as centre. If, from A, we draw a sphere with radius AP = CF, we get the position of this spherical wave when the disturbance originally at C has reached the reflecting surface. Similarly, DE being parallel to ABC, if we draw from D as centre a sphere with radius DQ = EF we get the corresponding position of the spherical wave which is originated at the point D when the wave-front reaches it. All such spheres touch a plane surface, PQF, which is, therefore, the wave-front after reflection. Thus we see that a plane wave remains a plane wave after reflection.

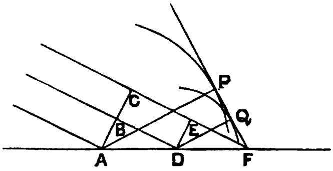


FIG. 74.

But, further, AP = CF, and AF is common to the two right-angled triangles ACF and APF. Therefore the angles CAF and PFA are equal, that is, the reflected wave-front has the same angle of inclination to the reflecting surface as the incident wave-front has. And this—since the rays are perpendicular to the wave-fronts—gives the known law of reflection.

162. *Laws of Refraction.*—A ray of light, on passing from one medium into another of different density, is in general bent from its original direction, and is said to be refracted. The angle which the refracted ray makes with the normal is called the angle of refraction.

The refracted and incident rays lie in one plane with the normal to the surface and the sine of the angle of incidence bears a constant ratio to the sine of the angle of refraction.

The angles of incidence and refraction (Fig. 75) being denoted by the letters i and r respectively, the above law may be written

$$\sin i = \mu \sin r.$$

The constant, μ, is called the *index of refraction.* When the refraction takes place from a less dense into a more dense medium, μ is generally greater than unity; and, conversely, μ is usually less than unity when the light passes from a more dense into a less dense medium.

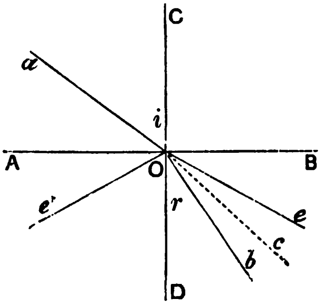


FIG. 75.

Gladstone and Dale found that the relation $(\mu - 1) = k\rho$, where k is constant and ρ is the density of the medium whose refractive index is μ, holds as a close approximation. Now, ρ is equal to the product of the molecular mass into the number of molecules per unit volume. Hence, the refractive index of space free from matter being supposed to be unity, so that the space between molecules contributes nothing to the refraction, we see that this relation means that each molecule has a definite refractive power. Its value can be calculated in the case of some organic compounds when the chemical constitution of the substance is known. This is only one of many examples in which molecular properties are found to be constant (§§ 240, 244, 250). Jamin thought that, in water, μ^2-1 was proportional to ρ; but Mascart showed that Jamin's observations accorded more closely with the law of Gladstone and Dale; and this was confirmed by Quincke. Solids and liquids are so incompressible, and the refractive indices of gases differ so little from unity, that very careful experimental work is necessary in order to determine the superiority of the latter expression. Jamin's law is a consequence of the corpuscular theory.

The following table gives rough values of the refractive indices of a few solids and liquids at ordinary temperatures, and of a few gases at 0° C., yellow light being used:

Diamond ...	... 2·470	Ether ...	... 1·356
Dense flint-glass ...	1·650	Water ...	... 1·332
Light flint-glass ...	1·540	Carbonic oxide ...	1·000454
Rock-salt ...	... 1·544	Air ...	... 1·000292
Carbon bisulphide	1·634	Oxygen	... 1·000271
Alcohol ...	... 1·365	Hydrogen	... 1·000139

Kundt has determined the refractive indices of some metals by using red light with wedges so thin as to be transparent. His results varied from 0·27 in the case of silver, to 2·61 in the case of bismuth. Rubens, by a theoretical deduction from observed values of the reflecting power of the metals, has obtained concordant results. Ayrton and Perry found 1·66 for the index of ebonite with light of large wave-length. Shea has recently repeated Kundt's observations with results in fair agreement.

A very direct method of determining the refractive index of a liquid was given by Newton. A rectangular vessel, with a plane glass bottom, contained the liquid upon the surface of which a ray of light fell obliquely. The vessel was tilted till the ray emerged perpendicularly to the bottom, when its inclination to the vertical gave the angle of refraction. Other methods will be described subsequently (see §§ 167, 190, 197).

The law of refraction was first given by Snell about 1621.

The intensity of the refracted ray depends upon the angle of incidence. It is greatest when the incident ray is perpendicular to the refracting surface, and diminishes as the angle increases. In the case of refraction into a denser medium, minimum intensity is attained at grazing incidence: in the case of refraction into a rarer medium, the minimum is reached when the angle of incidence is less than 90°; and at higher angles of incidence no refraction occurs.

163. *Refraction through a Plane Surface.*—Let a ray, aO (Fig. 75) fall upon the plane surface, AB, of a medium, the refractive index of which is μ—that of the first medium being taken as unity. If we assume μ to be greater than unity, the path of the ray in the second medium will be a line Ob, such that $\sin a\text{OC} = \mu \sin b\text{OD}$, CD being perpendicular to AB at the point of incidence O. Also light travelling in the direction bO will emerge into the first medium in the direction Oa, which is such that $\sin b\text{OD} = 1/\mu \sin a\text{OC}$, the refractive index of the first medium with reference to the second being the reciprocal of that of the second with reference to the first.

Now suppose $a\text{OC} = 90°$. The light will enter the second medium in a direction Oc, such that $\sin c\text{OD} = 1/\mu$; and, conversely, light travelling in the direction cO will emerge into the first medium so as just to graze along the surface. Light passing in the direction eO, where $e\text{OD} > c\text{OD}$, cannot enter the first medium at all, but will proceed in the direction oe', in accordance with the ordinary law of reflection. (This proves the concluding remark of last section.) The ray eo is said to suffer *Total Reflection*, and the limiting angle cOD is called the *Critical Angle*.

The refractive index of water, relatively to that of air, is about 4/3 for ordinary yellow light. Hence an eye placed underneath the surface of water will see objects above the surface by means of rays which are crowded into a cone, the sine of the semi-vertical angle of which is 3/4.

The angle of deviation of a ray from its original direction by a single refraction is $i-r$, i and r being respectively the angles of incidence and refraction.

From O (Fig. 76) as centre describe two circles APB and CQD, with radii equal to unity and μ respectively. Let $\text{COQ} = r$, and draw QPN parallel to CO. Then $\text{ON} = \sin \text{OPN} = \mu \sin \text{OQN} = \mu \sin r$, and so $\text{OPN} = i$.

Now suppose that P moves out from A through a given infinitesimal distance into a position P′, and let OQ′ be then turned back to coincide with OQ. P′ will then take a position p such that pQ is greater than PQ. And the difference between pQ and PQ becomes greater and greater as P moves out from A through successive equal distances, for the circle AB has greater curvature than the circle CD. But the difference between pQ and PQ is the growth of PQ during a given very small increase of i.

Hence PQ grows at an increasing rate as i increases uniformly. But PQ subtends at O the angle $i-r$ and becomes more and more

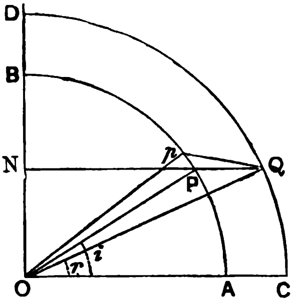


FIG. 76.

nearly perpendicular to OP and OQ as P moves out from A. Therefore *$i-r$ increases faster and faster as i increases uniformly.*

In the above process $i-r$ has increased by the amount Pp, and $PQO=r$ has increased by the amount $pQP = Pp$ NP/PQ. Hence $r-(i-r)=2r-i$ has increased by the angle Pp (NP/PQ - 1). This vanishes when NP = PQ. This implies AC < OA, that is, $\mu < 2$. Again, the increase of $2r-i$ is positive or negative according as NP > or < PQ. Hence it is always positive until the limiting angle is reached, after which PQ still increases, while NP diminishes, so that $2r-i$ diminishes continuously. In other words, the condition NP = PQ or $2 \cos i = \mu \cos r$ (since $NQ = \mu \cos r$) indicates a *maximum* value of $2r-i$.

This condition, combined with the condition $\sin i = \mu \sin r$, enables us to calculate the angle at which this maximum occurs.

Similarly, the angle $3r-i$ increases under the same conditions by the amount Pp (2NP/PQ − 1); and the condition for a stationary value is 2NP = PQ or $3 \cos i = \mu \cos r$, which implies $\mu < 3$. As in the former case, the condition indicates a maximum.. From this condition, along with $\sin i = \mu \sin r$, we can calculate the value of i which corresponds to the maximum. In each case the value of the angle depends only on the value of the refractive index.

These results are of importance in the discussion of the primary and secondary rainbows.

164. *Refraction Caustics.*—Let OPo' (Fig. 77) represent an axial section of an infinitesimal circular cone of rays diverging from a point P placed in a dense medium. The section, by the

plane of the paper, of the cone *a'pa*, by which the object P is actually seen by an eye situated in the rare medium, has its vertex at a different point, *p*, which is nearer the surface than P is.

The amount by which an angle must increase in order to produce a given increase in the value of its sine becomes greater as the angle is greater. Hence, if OP and *o'*P, while containing an angle of constant magnitude, become less inclined to the common surface of the two media, the angle *apa'* becomes greater and greater as O moves away from Q the foot of the perpendicular from P upon the surface. Thus, as O moves out from Q, O*p* separates out from OP at an increasing rate, and its length becomes smaller at an increasing rate. The result is that *p* always lies on that side of PQ on which O lies, and constantly approaches O as O recedes from Q. In particular, O and *o'* being indefinitely near to each other, *p* ultimately coincides with O when O*a* just grazes the surface. No more than two finitely distinct rays in any given vertical plane section can intersect in one point, *p*; but, when the cone is very small, all the rays in a given plane section pass approximately through one point.

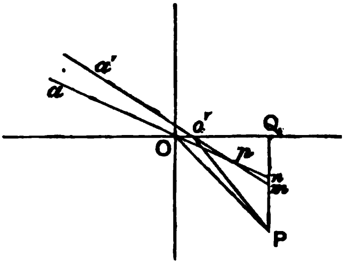


FIG. 77.

We have next to consider the lateral divergence of the beam of light. After refraction, a ray remains in the same plane, passing through the normal, which contained it before refraction. Hence, laterally, the rays diverge from points on PQ. And thus, since the cone has some lateral thickness at the point *p*, we see that the emergent light appears to pass, at *p*, approximately through a small line, the length of which is perpendicular to PQ.

This is the primary focal line.

The secondary focal line is the intersection of the refracted cone with a plane drawn perpendicular to its axial line and passing through the point in which the axial line cuts the line PQ.

At perpendicular incidence, $Op = On = OP/\mu$; for $OPQ = r$ and

$OnQ = i$ and $\sin i/\sin r = OP/On$. Therefore the light appears to diverge from a point which is closer to the surface than the actual point in the ratio of 1 to μ. For example, the depth of water appears to be only three-fourths of its true depth.

The existence of the primary focal line implies the existence of a caustic surface. We may map out (Fig. 78) successive points,

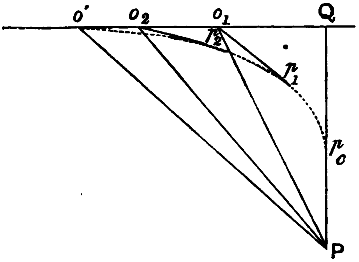


FIG. 78.

p_1, p_2, etc., corresponding to successive positions of the point o, by repeating the construction of Fig. 77. We thus obtain the caustic curve $p_0p_1p_2o'$, which touches QP at p_0 and Qo' at o'. The line Po' indicates the limiting position beyond which total reflection takes place. If the caustic curve be rotated around the line PQ, the caustic surface will be described.

165. *Refraction through Parallel Layers.*—A ray of light, which is bent out of its original direction on entering a refracting layer (Fig. 79) with parallel plane sides, is, on re-entering the original medium on the far side of the layer, bent back again into its first

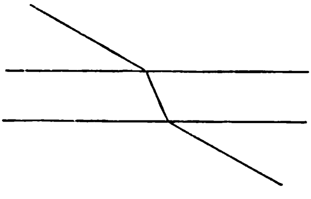


FIG. 79.

direction. This follows at once from the symmetry of the arrangement.

If an object is looked at perpendicularly through such a layer of thickness t, its distance from the eye is apparently diminished by the amount $t(\mu-1)/\mu$. For, if d be the distance of the object from the layer, an eye situated just inside the layer would see it as if it were at a distance $d\mu$ from the layer. Therefore, when the eye is just at the far side of the layer, but still inside, the apparent distance of the object is $t+d\mu$. And, if the eye be now just outside the layer, this distance is decreased in the ratio of unity to μ, which proves the statement.

If a ray of light passes through three such layers the total deviation of the ray from its original direction is the same as if the intermediate layer were absent. For, μ_1 and μ_2 being the refractive

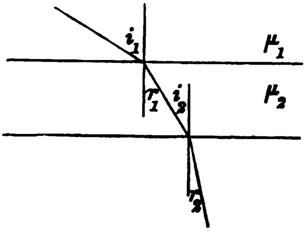


FIG. 80.

indices of the first and second media with reference to the second and third respectively, we have (see Fig. 80)

$$\frac{\sin i_1}{\sin r_2}=\frac{\sin i_1}{\sin r_1}\cdot\frac{\sin r_1}{\sin r}=\frac{\sin i_1}{\sin r_1}\cdot\frac{\sin i_2}{\sin r_2}=\mu_1\mu_2.$$

But, since μ_2 is the index of the second medium expressed in terms of that of the third as the unit, while μ_1 is the index of the first expressed in terms of that of the second as the unit, $\mu_1\mu_2$ is the index of the first medium expressed in terms of that of the third.

It follows that no effect is produced by any number of intermediate layers. For this reason it is possible to calculate the refraction error in the apparent altitude of stars, etc.

166. *Mirage.*—When a beam of light passes non-perpendicularly through a medium composed of parallel layers of continuously varying density, the direction of the beam constantly alters. Let us suppose that, over the surface of the ground AB (Fig. 81) there exists a stratum of air the density of which continually increases with its distance from the surface. We may suppose also that, above the plane *ab*, the density remains uniform. An object, *nm*, will be seen by an eye situated at *p* by means of rays which enter the non-homogeneous medium and are then bent upwards until, re-entering the uniform medium, they pass straight to *p*. The object

appears to be in the direction of *m′n′*, and is obviously inverted since the rays from the upper and lower extremities cross each other before reaching the eye. But the object may also be seen directly through the medium above *ab*, and so the impression of an object with an inverted reflection is produced. This is the case of the ordinary mirage of the desert.

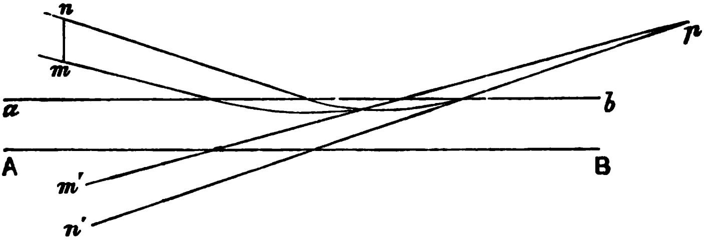


FIG. 81.

In Fig. 82 AB again represents the surface of the ground, and the density of the air is suppose to increase from below upwards. A direct mirage is thus produced.

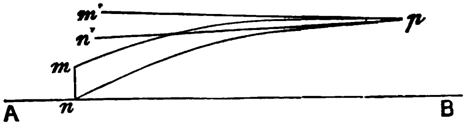


FIG. 82.

In Fig. 83 the air is supposed to be of uniform density between AB and *ab*, but is supposed to diminish continuously in density above *ab*. This condition obviously gives rise to an inverted image.

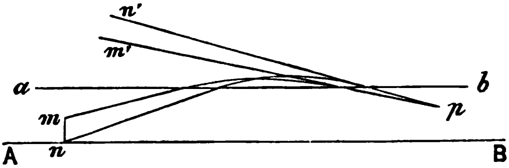


FIG. 83.

167. *Prisms.* — When the two plane bounding surfaces of a medium are not parallel, the emergent ray is no longer parallel to the incident ray. Such an arrangement constitutes a prism.

Let ABC (Fig. 84) represent the section, by the plane of the paper, of a prism of some dense medium, such as glass. Let the edge, B, of the prism be perpendicular to the plane of the paper, and let *abcd* be a ray which makes an angle i with the (inward) normal to the face AB and an angle i' with the (outward) normal to the face BC. Let r and r' be the corresponding angles of refraction.

The deviation of the ray from its original direction by the first refraction is $i-r$; and at the second surface this deviation is farther increased by the amount $i'-r'$. But (§ 168) the alteration of $i-r$, consequent on a given alteration of i, is greater and greater the larger i or r is; and any increase or decrease of r involves an equal decrease or increase of r'. Hence, if i be greater than i', a given diminution of i produces a diminution of $i-r$, which is greater than the simultaneous increase of $i'-r'$. Hence the ray has *minimum deviation* when $i=i'$, that is, when the ray passes through the prism so as to make $Bb=Bc$.

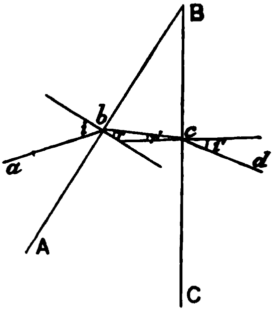


FIG. 84.

When the angle of the prism is less than the angle whose sine is $1/\mu$ the deviation $i'-r'$ may be in the opposite direction to the deviation $i-r$, but it is easy to see that the above result is true in all cases.

The total deviation, δ, is, in the standard case above, $i'+i-(r'+r)$, which, in the minimum position, becomes $2(i-r)=2i-\alpha$, α being the angle of the prism. Hence we can express i and r, and therefore the refractive index of the substance of which the prism is composed, in terms of the minimum deviation and the angle of the prism.

This affords a ready means of determining the refractive index of the substance of the prism; and, if the prism be hollow and have its sides made of glass plates of uniform thickness, the refractive index of any liquid or gas placed in the hollow may be found. Another simple method of determining refractive indices by means of a prism was given by Descartes. A beam of light was allowed to fall perpendicularly on one face of a prism so that it suffered no change of direction until it emerged at the other side. The direction of the light being now supposed to be reversed, the total deviation gives the angle of incidence, while the angle of refraction is equal to the known angle of the prism.

168. *Refraction through Spherical Surfaces: Lenses.*—Let QAB (Fig. 85) represent a part of a spherical surface the centre of which is at O; and let the refractive index of the substance on the convex side of the surface be μ. Let rays PA, PB, diverging from P, meet the surface in the points A and B respectively. Suppose that Ap, Bp are the backward prolongations of the refracted rays —μ being assumed to be greater than unity. Join PO and continue the line to meet the surface in Q.

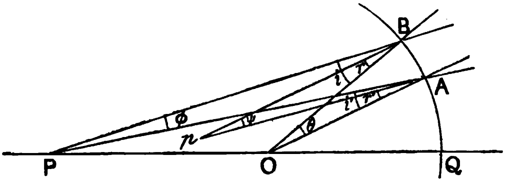


FIG. 85.

If the angles be denoted as in the figure we get at once, from the equality of vertical angles, $i' - i = \phi - \theta$, and $r' - r = \psi - \theta$. Hence, when the angles are so small that we may write $\sin i = i$, $\cos i = 1$, etc., we have (§ 163) $i' - i = \mu\,(r' - r)$, and therefore

$$(\mu - 1)\,\theta = \mu\psi - \phi \;.\;.\;.\;.\;.\;.\; (1).$$

When A coincides with Q and AB is very small this becomes

$$\frac{\mu - 1}{\mathrm{OQ}} = \frac{\mu}{p\mathrm{Q}} - \frac{1}{\mathrm{PQ}}.$$

When $\mu = -1$, this gives the first formula of § 158.

Now let the rays (Fig. 86), apparently diverging from the point p, fall upon the second spherical surface of the dense medium at the points A′, B′. Let O′ be the centre of the new surface, and let O′p meet it at the point Q′. From the diagram we get $i' - i = \psi - \theta'$ and $r' - r = \psi' - \theta'$. Hence, when A′B′ is very small, $\psi' - \theta' = \mu(\psi - \theta')$, or

$$(\mu - 1)\,\theta' = \mu\psi - \psi' \;.\;.\;.\;.\;.\;.\;.\;.\; (2)$$

[Of course (2) might have been deduced from (1) by the substitution of $1/\mu$ for μ, θ' for θ, ψ' for ψ, and ψ for ϕ.]

From (1) and (2) we get

$$(\mu-1)(\theta-\theta')=\psi'-\phi \ldots\ldots (3)$$

When A (Fig. 85) coincides with Q, A′ (Fig. 86) coincides with

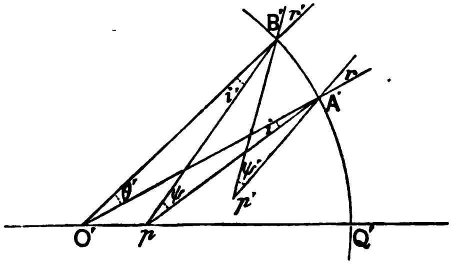


FIG. 86.

Q′; and so, when the angles are sufficiently small, we may write (3) in the form

$$(\mu-1)\left(\frac{1}{r}-\frac{1}{s}\right)=\frac{1}{v}-\frac{1}{u} \ldots\ldots (4)$$

r and s being the radii of the two spherical surfaces, while v and u are respectively the distances of p' from Q′, and of P from Q.

The quantity on the left-hand side of equation (4) is constant. We may therefore write (4) in the form

$$\frac{1}{f}=\frac{1}{v}-\frac{1}{u} \ldots\ldots\ldots\ldots (5)$$

where f is a constant, called the *principal focal distance.* When u is infinite, that is, when the incident rays are parallel, we get $f=v$; in other words, f is the distance from Q′ of the point from which originally parallel rays seem to diverge after passing through the medium. This point is called the *principal focus.* P and p' are called *conjugate foci.*

A portion of a medium bounded by two spherical surfaces is termed a *lens.*

In the case just discussed we have taken $s>r$, and have assumed that the concave sides of the spherical surfaces are directed towards P. All lines drawn in the direction PQ are regarded as positive. Lines drawn in the direction of QP are therefore negative.

Equation (4) applies to all lenses, if we remember that u and v are taken as positive or negative according to the direction in which they are drawn. The quantity f is negative (1), when s is positive but less than r, (2) when r is infinite and s is positive, (3) when r is negative and s is positive.

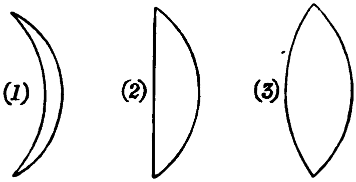


FIG. 87.

In such cases the principal focus and the source of the parallel rays are on opposite sides of the lens, which is necessarily thickest at the middle (Fig. 87).

But f is positive (1) when r and s are both positive and $s > r$, (2) when s is infinite and r is positive, (3) when s is negative and r is positive.

In the three latter cases the principal focus and the source of the parallel rays are on the same side of the lens, which is now thinnest (Fig. 88) at the middle. Such a lens diminishes the convergence or increases the divergence of the incident rays, while a lens of the previous type diminishes the divergence or increases the convergence of the incident rays. Equation (3) makes this evident at once.

The formula (4) is true of a given small cone of rays which falls perpendicularly upon the lens at its central part, and it shows that, after passing through the lens, all such rays appear to diverge from, or are converged to, another perfectly definite point. We shall see in next section that the same result holds in the case of a small obliquely-incident pencil of rays.

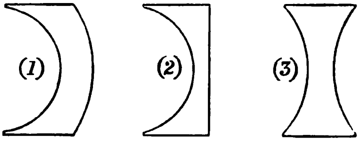


FIG. 88.

169. *Lenses: Oblique Refraction.*—Let DCBA (Fig. 89) represent a ray, which, entering the lens CBQQ′ at C, emerges at B and cuts the line OQ in A. (The letters O, O′, Q, Q′ have the same meanings as in last section.) Let the ray be such that OB and O′C are parallel. Since BC makes equal angles with OB and O′C, it

follows that AB and CD make equal angles with them. Therefore AB and CD are parallel.

Let CB be continued to meet OQ in R. We have OR/OB= O′R/O′C, and so the lengths of OR and O′R bear a constant ratio to each other. Therefore R is a fixed point, and is called the *centre of the lens.*

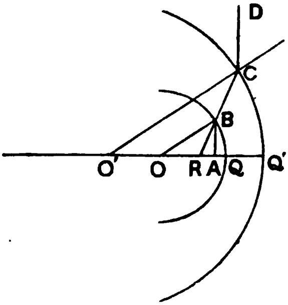


FIG. 89.

Thus, no ray, whose path in the lens passes through the centre, is deviated from its original direction by its passage through the lens.

It is easy to find an expression for the distance RQ in terms of the thickness of the lens and its radii of curvature. When the lens is very thin, R practically coincides with Q or Q′, or any intermediate point; and the two portions, on either side of the lens, of a ray passing through the centre, are practically in one straight line.

We shall hereafter assume that we are dealing with thin lenses only.

From the results of § 164, we see that, in order to converge to a definite point in a dense medium, an inclined pencil of rays incident upon the plane surface of that medium must have its convergence in the plane of incidence different from its lateral convergence. Conversely, a uniformly convergent incident pencil will not converge to a single point, but will pass through two focal lines unless the incidence is normal. In the case considered in § 164, since the caustic surface touches QP (Fig. 78) at the point corresponding to normal incidence, we see that, as *o* moves out from Q, the distance between the primary and secondary focal lines increases at first at a very slow rate; and the same thing holds true in the converse problem which we are now considering. Hence, the rays still seem to converge to one point when the angle of refraction is appreciable though small. And, since the angle of incidence is greater than the angle of refraction, the statement holds still more closely when made with reference to the former angle instead of to the latter. Also, when the rays, having passed through the medium, emerge at a parallel plane surface, the action is simply reversed, so that the rays still seem to move to or from a single point. Now, in a very thin lens, the surface differs little from a plane. We therefore conclude that a given, sufficiently small and not too oblique, pencil of rays will practically, after refraction through a thin lens, be brought to a focus at, or appear to diverge from, one definite point; and this point must be situated on the line drawn from the vertex of the incident pencil through the centre of the lens.

We are now in a position to investigate the production of images by means of lenses.

170. *Formation of Images by Lenses.*—Let AB (Fig. 90) be a thin lens, of which C is the centre and CF is the principal focal length. If MN be an object which is situated at a distance from the lens greater than CF, rays diverging from N will be brought to a focus at a point n such that $1/CF = 1/CN + 1/Cn$.

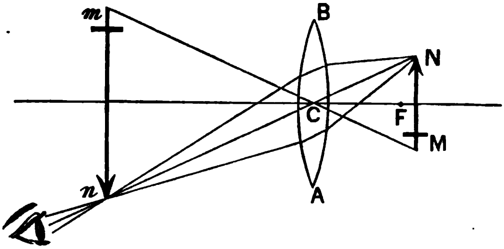


FIG. 90.

Similarly $1/CF = 1/CM + 1/Cm$, and so on.

Thus an inverted real image of MN is formed at *mn*, and the rays diverging from that image may be examined by an eye which is situated at the distance of about ten inches from it—that being the usual distance for most distinct vision.

On the other hand, if MN be slightly nearer (Fig. 91) to the lens than the principal focus, the lens is only able to diminish the divergence of the incident rays, and the erect and virtual image of MN is situated at a greater distance from the image than the object is. For the purpose of correct vision as regards an eye placed close behind the lens, this distance must be about ten inches.

The *magnifying power* of the lens is the ratio of *mn* to MN, and is therefore approximately equal to $10/f$, where f is the principal focal length expressed in inches.

The object-glass of the ordinary *astronomical telescope* acts in

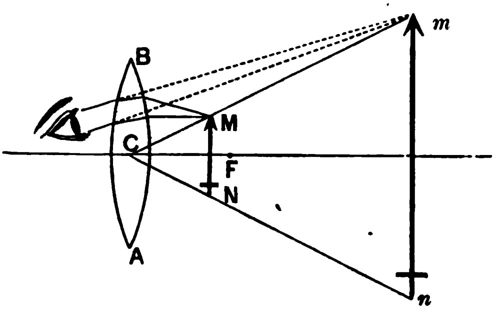


FIG. 91.

the manner first described above. The practically parallel rays from a very distant object, MN (Fig. 92) are converged to the principal focus of the object-glass, and so an inverted image, *mn*, is formed.

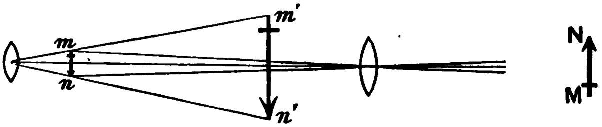


FIG. 92.

The eye-glass is placed at a distance from *mn* which is slightly less than its principal focal length, and so forms a magnified image, *m'n'*, still inverted, at a distance of ten inches from the eye.

The magnifying power of the telescope is the ratio of the angle which *mn* subtends at the eye-glass to the angle which it subtends at the object-glass. It is, therefore, approximately equal to the ratio of the focal length of the object-glass to that of the eye-glass.

The arrangement of lenses in the *compound microscope* is essentially the same. The object is placed at a distance from the object-glass which is slightly greater than its principal focal length, and so a magnified inverted image is formed, and is further magnified by the eye-glass.

171. *Dispersion: Aberration.*—In all the preceding sections it was assumed that we were dealing with light of one definite kind

or colour alone. But rays of light of different colours are differently refracted by any given substance; and we must, therefore, with this in view, reconsider briefly the action of prisms and lenses.

Let a ray of white light, *ab* (Fig. 93), fall upon a prism, ABC, in a direction perpendicular to its edge. The single ray of white light will, on entering the prism, be broken up into a series of coloured rays, *bc*, *bc'*, etc. The red rays are least deviated, and the blue or violet rays are most deviated, from their original direction.

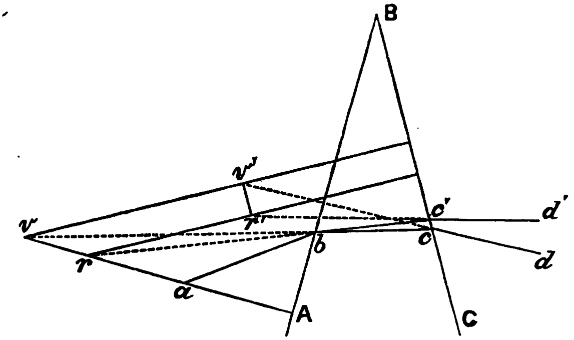


FIG. 93.

Let *bc* and *bc'* represent respectively a violet, and a red, ray ; and let these lines meet the perpendicular from *a* on AB in the points *r* and *v*. So long as the angle of incidence is very small, an eye placed in the substance of the prism will see, no white point *a*, but a coloured line, *rv*, red at the end nearest to, and violet at the end farthest from, the prism—the ratios of *r*A and *v*A to *a*A being respectively the ratios which the refractive indices of the substance of the prism for red light and violet light bear to unity (see § 164).

After emergence from the face BC, the two rays considered will take the directions *cd*, *c'd'*. Drop the perpendiculars *rr'* and *vv'* on BC, and let *dc* meet *vv'* at the point *v'*, while *d'c'* meets *rr'* at the point *r'*. The angles which the emergent rays make with BC being large (which necessitates ABC being small), we see that an eye placed in the same medium as the point *a*, but on the opposite side of the prism from it, perceives, instead of *a*, a coloured line, *r'v'* ; for the points *r'*, *v'*, are situated respectively with regard to the face BC and the points *r*, *v*, in the same way that *r*, *v*, are situated with regard to the face AB and the point *a*.

Let us suppose now that *a* represents the section, by the plane of the paper, of a luminous white line, which is parallel to the edge of the prism. In this case $r'v'$ represents a coloured band, which is called a *spectrum*. This separation of the constituents of white light by refraction is termed dispersion. The measure of the dispersion produced by any given substance is the difference of the refractive indices of that substance for the extreme rays of the visible spectrum.

Now let P (Fig. 94) be a luminous white point from which diverging rays fall upon a lens AB. The refractive index of the substance of the lens for violet rays being greater than its refractive index for red rays, the violet rays will be brought to a focus at a point, p, which is nearer the lens than the point p', to which the red rays converge. The light in the regions BpA and B$'p'$A$'$ is nearly colourless, but, on the whole, it is somewhat violet in the former and reddish in the latter. In the region immediately outside BpA, the light is red, while, in the region immediately outside B$'p'$A$'$, it is violet.

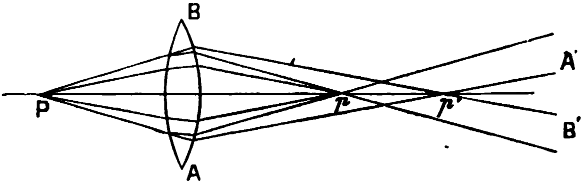


FIG. 94.

A lens is said to produce *aberration* when it fails to bring all rays diverging from one point to a focus at another point. The aberration is called *chromatic aberration* when it is due to dispersion: it is called *spherical aberration* when it is due to the spherical form of the faces of the lens. The former increases in proportion to the angle of incidence, while the latter, as we saw in § 169, increases at first at a rate which is vanishingly small in comparison. So long as the lens is sufficiently thin, and the incidence is sufficiently direct, both are negligable.

172. *Dispersion: Achromatism.*—It is possible to get nearly rid of dispersion while refraction remains, and thus we can obtain a practically achromatic lens. This result can be obtained, since some highly refracting substances produce comparatively small dispersion, while some substances of low refracting power produce comparatively large dispersion—comparatively, that is, to their refraction.

Let us suppose that two prisms, of equal angles, but of different substances, produce dispersions d_1, d_2 respectively, and let their refractive indices for the extreme red light of the spectrum be μ_1, μ_2 respectively. If now we alter the angle of the prism, the dispersion of which is d_1, in the ratio d_2/d_1, and form a compound prism of the two with their edges turned in opposite directions, we shall have a prism which will not produce dispersion of the red and violet rays; but refraction will still take place unless $d_1/\mu_1 = d_2/\mu_2$.

Similarly, we may construct a compound lens for the purpose of avoiding chromatic aberration.

By no pair of substances yet found, however, can we produce complete achromatism. If a lens is completely achromatic for two definite kinds of light it will not be so for any other pair. For if given dispersion is produced by one prism, between a series of pairs of definite kinds of light, equal dispersion will not be produced, by any other prism, between more than one of these pairs. This is known as the *Irrationality of Dispersion.*

Compound lenses, made up of three constituent lenses, can produce closer approximation to achromatism than can a compound lens made up of only two.

If that one of the two equal prisms, referred to above, whose refractive index is μ_1, has its angle altered in the ratio of μ_2 to μ_1, the combination of the two prisms with their edges reversed will give no refraction of the red ray, but will give dispersion unless the ratio of the refractive indices of the two prisms for each ray of the spectrum is constant. This principle is used in the construction of direct vision spectroscopes (§ 178).

173. *Rainbows : Halos.*—We are now in a position to discuss the formation of rainbows and halos.

Let AB (Fig. 95) represent a ray of yellow light which, falling on a drop of water at B, is refracted to C. After reflection at C the ray emerges at D. The whole figure is symmetrical about the line drawn from C through the centre of the drop. The semi-angle between AB and ED is obviously $2r - i$, i and r being respectively the angles of incidence and refraction.

The refractive index of water for yellow light is about 1·336; and so, by the calculation referred to in § 163, we find that the semi-angle, corresponding to the maximum value of $2r - i$, is equal to 21° 1′ nearly. But the existence of a maximum means that the rays are crowded closely together in the immediate neighbourhood of the maximum angle, and so an eye situated at E will see comparatively strong yellow light in the direction ED.

Now consider AB to be a ray of white light. This becomes dispersed on refraction, and a blue ray (say) being more refracted, will suffer reflection at a point C′ such that the angle $2r-i$ is smaller than before. Hence the maximum value of the angle $2r-i$ becomes smaller as μ increases.

From E draw EP parallel to AB, and let a very large number of drops be situated symmetrically around EP in the direction of the sun's rays. (Of course, in the figure, the size of the drop is immensely exaggerated relatively to other magnitudes.) The eye at E will see a circle of yellow light, of radius 21° 1′, the centre of which is situated on the line EP. Inside this a circle of blue light will be seen, and, outside it, a red circle will appear—the colours of the spectrum succeeding each other, in order of decreasing refrangibility, from within outwards. This constitutes the explanation, by geometrical optics, of the primary rainbow. In the actual bow the colours are, of course, impure. For the light proceeding from each point of the sun's disc gives rise to a separate bow, and the bow which we see results from the superposition of all these distinct bows.

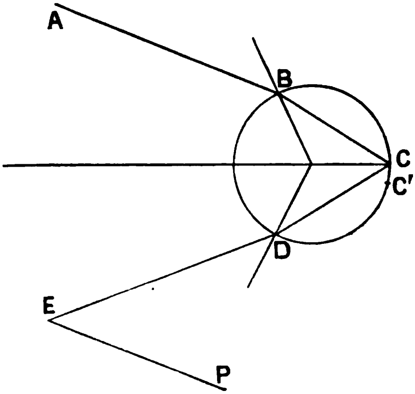


FIG. 95.

If the ray AB (Fig. 96) suffers two reflections in the interior of the drop, the emergent ray EP makes an angle with it which is double of $\pi/2-(3r-i)$. But $3r-i$ reaches a maximum when i has a certain value (referred to in § 163). Hence the angle BPE is a *minimum* when i has this value. For yellow light, falling on a drop of water, the vertical angle is about 50° 58′.

A ray of higher refrangibility will be reflected at points C′, D′, such that the perpendicular from O on C′D′ will intersect AB at an angle which is larger than BPO. Hence the minimum value of the angle $\pi/2-(3r-i)$ increases as μ increases. And so, in the secondary rainbow, which is due to two such internal reflections, the colours succeed each other, in order of increasing refrangibility, from within outwards.

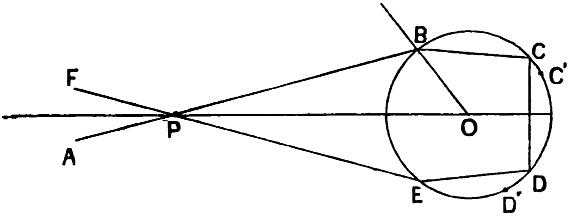


FIG. 96.

Also, since the primary bow corresponds to a *maximum* value of the angle which the emergent ray makes with the incident ray, while the secondary bow corresponds to a *minimum* value of this angle, we see that the space between the bows is devoid of light due to rays which have suffered one or two reflections inside the drops, while there is some such illumination in the region inside the first bow and also in that outside the second.

The bows which are due to more than two internal reflections are too feeble to be visible.

Halos are due to the refraction of light through ice-crystals. The rays are most densely crowded in the direction of minimum deviation. The red rays are least deviated, and therefore appear always on the interior portions of halos. The size of a halo depends upon the effective angle of the ice-crystal to refraction through which it is due. *Parhelia* and *paraselenæ* are simply exceptionally bright portions of halos where two or more intersect.

Colourless halos are produced by the reflection of light from the plane surfaces of the crystals.

174. The general phenomena of reflection and refraction receive a ready explanation whether on the corpuscular or on the undulatory theory of light.

Let AB (Fig. 97) represent the bounding surface between two refracting media. Let $p\ q\ r\ s\ t$ represent the path of a corpuscle. During the rectilinear motion from p to q the corpuscle experiences no resultant attraction in any direction. When it reaches the distance from AB indicated by the line ab, the greater (say) attraction of the medium on the other side of AB preponderates and the path becomes concave towards the surface of separation.

This continues until a point, *s*, at which the resultant attraction again becomes zero, is reached. The rest of the path *st* is therefore straight, and is inclined at a less angle to the normal to the surface AB than is the part *pq*.

Even if the refracting surface is not actually plane, it is yet practically plane, in all cases of finite curvature, so far as the present reasoning is concerned, for the portion *qs* of the path of the corpuscle is excessively small.

Now (*cf.* § 30) the square of the resolved part of the speed of the corpuscle along the normal increases by a constant amount in passing from *ab* to *a'b'*, while the resolved part along the surface remains constant, and hence the total speed, v', of the particle in the second medium bears a constant ratio to its total speed, v, in the first. Let, as usual, i and r denote respectively the angles which *pq* and *st* make with the normal. We have $v \sin i = v' \sin r$, which is identical with $\sin i = \mu \sin r$, where μ is equal to v'/v. But this is the known law of refraction.

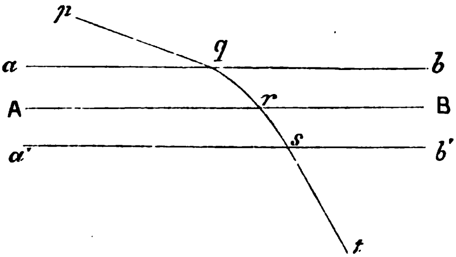


FIG. 97.

It follows necessarily that the speed of a corpuscle is greater in a dense medium than in a rare one.

175. Let ABC (Fig. 98) represent a plane wave-front, which travelling through the air in the direction indicated by the arrows, reaches the surface, ADF, of a refracting medium. From A, as centre, describe a sphere of radius, AP, such that light will travel over the distance AP in the refracting medium in the same time that it will describe the distance CF in the air. Similarly, from D describe a sphere of radius, DQ, such that DQ bears to EF (DE is parallel to ABC) the same ratio as the speed of light in the refracting medium bears to its speed in air. It is obvious from the construction that all such spheres touch a plane, PQF, which is therefore the wave-front after refraction.

Let us denote the speeds of light in the medium, and in air, respectively, by the letters v', v. Let i and r be respectively the angles of incidence and refraction. Then $CAF = i$, $AFP = r$, and $CF = AF \sin i$, $AP = AF \sin r$. But $CF/AP = v/v'$. Hence $\sin i = v/v' \sin r = \mu \sin r$, if $\mu = v/v'$; and so the known law of refraction is a consequence of the wave-theory of light.

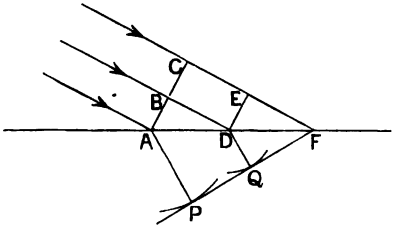


FIG. 98.

Observe that, in a dense medium in which μ is larger than unity, v' is necessarily less than v. Hence, on this theory, the speed of light must be less in a dense medium than it is in a rare one. This conclusion is in direct opposition to that derived from the principles of the corpuscular theory; and so we are furnished with a *crucial test* between the two theories. The result of experiment is (§ 153) entirely in favour of the undulatory theory. Consequently, we shall hereafter deal with the results of this theory alone.

It is easy to deduce from the above theory the fact that the time taken by light to travel from a given point in one medium to a given point in another is a minimum. Let PAQ (Fig. 99) be the actual path of light proceeding from P to Q, and let PBQ be a *very near* path. Draw AC and BD perpendicular to PB and AQ respectively. Then $CB/AD = \sin i/\sin r = v/v'$. That is, $CB/v = AD/v'$, or AD and CB are described in equal times. Therefore, and since practically $PA = PC$, $DQ = BQ$, the paths PAQ and PBQ are described in prac-

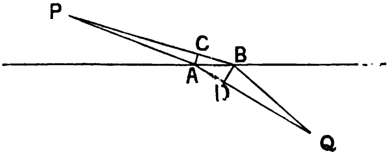


FIG. 99.

tically equal times. But, as B moves away uniformly from A, QA − QB increases at a diminishing rate, while PB − PA increases at an increasing rate. Hence the time of description of PAQ is less than the time of description of any conterminous path finitely distant from it on the same side as PBQ; and similar reasoning applies to the case of paths on the opposite side.

This law of least time was first given by Fermat. It is true in the case of reflection also (see § 157). The corpuscular theory gives $PAv + AQv' =$ a minimum. This sum is termed the *action*.

CHAPTER XV.

RADIATION AND ABSORPTION: SPECTRUM ANALYSIS. ANOMALOUS DISPERSION. FLUORESCENCE.

176. WE have already discussed the reflection (including scattering) and refraction of light at the common surface of two media. We have now to consider the absorption of light in its passage through material media, together with other associated phenomena.

According to the undulatory theory (which we now assume to be true, and of the truth of which we shall receive additional evidence as we proceed) light consists of waves propagated through a medium (called the *ether*) which fills space.

The particles of a body which is emitting radiation must therefore be in rapid vibratory motion, and must communicate their motion to the ether. The parts of the body, the vibrations of which are communicated to the ether, may be the molecules, or the constituent parts of the molecules, or the atoms.

When a bell is struck violently and frequently, the resulting sound is extremely complex and consists of notes, of various pitches, which may differ greatly from each other in intensity. The more violent the blows upon the bell become, and the more rapidly they are made, the more complex will the sound be. New forced vibrations appear, and the intensity of each of those previously existing is increased. It is only when the blows are excessively feeble and unfrequent that the fundamental tone is heard by itself.

Now, the molecules of a solid body, at a high temperature, are constantly colliding; and the vibrations induced in an atom by one collision do not die out before another collision is sustained. Hence the radiation given off by such a body consists of vibrations of many periods; and, as the temperature of the body becomes higher, vibrations of shorter and shorter period make their appearance, and the intensity of all previously existing vibrations becomes greater. Thus, if we examine the spectrum of a body the temperature of which is gradually raised, we may at first perceive no coloured radiation at all; but, as the temperature rises, first red light may appear, then yellow, and so on, until a complete continuous spectrum is seen, and the luminosity of every part gradually increases as the temperature rises. By gradually increasing the temperature of the carbon filament of an incandescent light, Weber found that the gray light which first appeared (§ 155) came from that part of the spectrum which contained most energy at high temperatures, and that the luminosity gradually spread from that part towards both ends of the spectrum as the temperature rose. This is in accordance with the above principles, and shows that, in this case, the radiation which was most powerfully emitted at the highest temperature reached was that which was most readily excited at the low temperature.

In the case of an ordinary gas, however, the atoms usually are sufficiently free from collisions to be able to vibrate in their own proper modes. Hence radiations of definite periods only will be emitted; and thus the spectrum of a gas is discontinuous, and consists of bright lines. It varies with the temperature and pressure. As the pressure is increased the lines broaden out, and the spectrum gradually becomes continuous, like that of a liquid body or of a solid.

We already know (§ 148) that a body which has a definite period of vibration, and which is at rest, may be set in vibration by the communication to it, through an intervening medium, of vibrations, of its own proper period, which are emitted by another body. Hence, if radiation travelling through the ether enters a material medium (solid, liquid, or gas), and if the natural period of oscillation of the molecules or atoms coincides with the period of some of the ethereal vibrations, the matter will be set in motion, and the energy of the radiation will be diminished.

This is the process which is termed *absorption*.

In many cases (if not really in most cases) the period of the induced vibration is longer than that of the ethereal vibration which induces it (§ 182). In almost all cases the absorbed energy is manifested by a rise of temperature. Lampblack is found to be the substance which is most efficient in absorbing the total radiation given out by a source. It is, therefore, employed in all instruments used for the absolute measurement of radiated energy. Of these, the radiometer has been already described (§ 129) and the thermopile and bolometer will be discussed subsequently.

Radiation from which certain wave-lengths are absent, such as that of ordinary gases, is termed *selective radiation;* and we speak similarly of *selective absorption* when radiation of one or more definite wave-lengths is absorbed while other radiation is transmitted.

In every transformation of energy, it is desirable to use the most efficient mechanism, *i.e.*, the mechanism which will most completely change the supplied energy into the required form. Thus, in the production of luminous radiation, it is desirable to have an arrangement which will give selective radiation of *light*—all other radiation being absent. But all our ordinary sources of light are very wasteful. Langley has shown that, in the radiation of a gas-flame, only 2·4 per cent. of the total energy is associated with waves which give rise to the sensation of light. In the radiation of the electric arc, 90 per cent. is unavailable, and in sunlight 65 per cent. is unavailable. On the other hand, Langley showed that all the radiation from the firefly affects the retina.

Selective absorption is the chief cause of the distinctive colours of bodies (§ 179). In some cases the absorption of light gives rise to chemical combination. Thus, chlorine and hydrogen combine under the action of light. In other cases chemical decomposition ensues, as when plants decompose carbonic acid gas. The decomposition of silver salts is the basis of photography. The radiation which is most effective photographically is non-luminous radiation of short wave-length—usually called 'ultra-violet radiation,' and sometimes 'actinic rays.' For this reason it is possible to photographically investigate the ultra-violet spectrum. Eder discovered that silver bromiodide could be made sensitive to luminous radiation over a great range by the use of some colouring matters. By such means Abney has succeeded in photographing ultra-red radiation. And Langley, by an ingenious employment of the bolometer, has automatically registered the ultra-red spectrum, over a still greater range, with an accuracy rivalling that of a powerful spectrometer.

The absorption of light by gases is strongly selective, just as the radiation is. The total absorption of light by the air is very feeble, yet its absorptive power for certain wave-lengths is very strong (§ 178). Tyndall showed that there are great differences in the absorptive powers of gases. He found, amongst others, the following values for the absorptions of various gases, at a pressure of the thirtieth part of an atmosphere, for heat from a source at low temperature, relatively to air.

Oxygen ...	... 1	Carbonic oxide ...	... 750
Hydrogen	... 1	Nitric oxide ...	... 1,590
Nitrogen	... 1	Hydrogen sulphide	... 2,100
Chlorine	... 60	Ammonia ...	... 7,260
Bromine	... 160	Sulphurous acid	... 8,800

177. *Equality of Emissivity and Absorptive Power.* The *Absorptive Power* of a body, under given conditions, for any definite radiation, is the fraction of the whole incident radiation of that kind which it absorbs. Now, a perfectly black body is one which absorbs all the incident radiation; so we might define the absorptive power for the given radiation as the ratio of the amount of it which the body absorbs to the amount of it which a black body would absorb.

The *Emissivity* of a body, at a given temperature, for any given radiation, is the ratio of the quantity of that radiation which it emits to the quantity of it which is emitted by a black body under the same conditions.

An extremely simple relation connects these quantities: *The Emissivity and Absorptive Power of a body, at a given temperature, for any radiation, are equal.*

The proof of this law was given by Stewart in 1858. The law is (see § 229) an extension of the statement, made by Prevost about a century ago, that the radiation emitted by a body depends solely upon the nature of the body and upon its temperature; and various experimental illustrations of it were known long before Stewart's proof was given. Fraunhofer, in 1817, showed that the double yellow line in the spectrum of burning sodium coincides closely with the dark double line D in the solar spectrum. Swan afterwards proved that the coincidence was extremely exact, and that a small amount of common salt was capable of giving the bright lines strongly. Brewster had shown that definite portions of the sun's light are absorbed in its passage through the earth's atmosphere, and that additional dark lines or bands could be produced in the solar spectrum by passing the light through nitrous acid gas. Foucault had, in 1849, pointed out that, while the electric arc emits (more freely than its other radiations) yellow light of two definite refrangibilities, the light from one carbon pole is robbed of these two kinds of radiation when it passes through the arc. Stokes also had, in 1850, explained this by the analogous properties of sounding bodies (§ 148). Stokes's explanation does not constitute a *proof* of the law of equality of radiating and absorbing powers unless it be assumed that the mechanism of radiation has analogies with elastic material systems. On the other hand, the completeness of the explanation in itself furnishes strong presumptive evidence in favour of the truth of the assumption. Stewart's proof is based solely on an experimental fact.

It is known, as an experimental result, that a number of bodies, at different temperatures, placed inside an enclosure which neither allows radiation to pass outwards from within it nor inwards from without it, will ultimately arrive at, *and maintain*, one common temperature. But this could not result unless each body emitted radiation at precisely the same rate as that at which it absorbed the radiation. This proves the law so far as radiation *as a whole* is concerned. No such enclosure as has been postulated exists in nature; but a polished reflecting surface of silver would form a sufficiently close experimental approximation: and, the more nearly the condition is satisfied, the more nearly does the result hold.

De la Provostaye and Desains had followed out Prevost's theory to this extent, pointing out that the sum of the radiating power and the reflective power of the enclosure must be constant, and that, since the reflected radiation is in general polarised (Chap. XVII.) the emitted radiation must be oppositely polarised to the same extent—a conclusion which they verified experimentally.

The radiation inside the enclosure must be that of a black body at the same temperature, for any one of the bodies might be a black one. Let us suppose that one of the bodies absorbs one definite radiation only, and allows all others to pass freely through it. Solutions of didymium salts approximately possess this property. This body must emit the same kind of radiation as it absorbs, and that to precisely the same extent; otherwise its temperature would vary. This proves the law as stated for any definite radiation.

Many experimental illustrations of the truth of the law were given by Stewart and by Kirchoff. Stewart's experimental investigations, given in 1858, referred to non-luminous radiation only. Early in 1860 he published the results of the extension of them to luminous radiation. In the end of 1859, Kirchoff, working quite independently, had published the results of similar investigations with regard to light. Thus, a piece of red glass, held in front of a fire, appears red because it absorbs the green and blue rays. If placed in the fire it becomes colourless when its temperature becomes equal to that of the fire—for it then still allows the red rays to pass through it, and, in addition, itself emits the rays which it absorbed. If taken out of the fire and held in a dark room it emits bluish-green light—precisely that which it absorbed.

Again, Stewart, and also Kirchoff, showed that a plate of tourmaline, cut parallel to the axis of the crystal, emits, when heated, light which is polarised perpendicularly to that which it allows to pass, that is, it emits the rays which it absorbs when cold.

178. *Spectrum Analysis.*—Kirchoff's investigations formed the foundation of the powerful method of research known as *spectrum analysis*.

In order to examine the luminous radiation emitted by a given body, we may place in front of the body a narrow vertical slit, which is situated at the principal focus of a convex lens. The light diverging from the slit is thus condensed into a parallel beam which is passed through a prism (usually of dense glass), and so gives rise to a spectrum. This spectrum is magnified by means of a telescope. Such an arrangement essentially constitutes the instrument called a *spectroscope* (or *spectrometer*, if a graduated circle and vernier are attached for the purpose of determining the angular positions of the telescope when it is directed towards different parts of the spectrum).

Let us suppose that we are examining the light emitted from a highly-heated lime-ball, and that this light, before falling on the slit, passes through a Bunsen flame in which metallic sodium is being vaporised. The lime-ball alone would give a continuous spectrum. The sodium-tinged flame alone would give a discontinuous spectrum consisting of two bright yellow or orange lines situated close together. The spectrum actually observed is continuous, but has two bright lines in the same position as those in the spectrum of the sodium flame.

If we vaporise the sodium in the flame of a spirit lamp instead of in a Bunsen flame, everything else remaining the same, a continuous spectrum, crossed by two *dark* lines in the positions of the former bright ones, will be seen. This was pointed out and explained by Kirchoff.

The cause lies in the difference of the temperatures of the Bunsen flame and the flame of the spirit lamp. A line will appear bright, or dark, according as the intensity of the radiation of that particular kind from a black body at the temperature of the flame exceeds, or falls short of, the intensity of the radiation of that kind which is emitted by the source. For, if R be the intensity of the given radiation as emitted from the source, while $R' = pR$ is the intensity of the light of that kind emitted from a black body at the temperature of the flame, and ρ is the radiating power (or absorptive power) of the flame for that radiation, the intensity of the given kind of light which reaches the eye is $R - \rho R + \rho p R = R[1 + \rho(p-1)]$. This quantity exceeds or falls short of, R, according as p is greater, or less than, unity. If the source were a black body, p could not exceed unity unless the temperature of the flame were greater than that of the source; but, the source not being a black body, p may exceed unity, although the temperature of the flame is below that of the source, *i.e.*, bright lines may be visible. If, however, the difference of temperature of the source and flame be sufficiently great, the lines will appear dark.

The spectrum of sunlight was found by Wollaston, in 1802, and by

Fraunhofer, in 1814, to exhibit a number of persistent dark bands. In accordance with the above principles we conclude that these lines are due to absorption. Some of them can be shown to be due to absorption in the earth's atmosphere, but the great majority are produced by absorption in the (comparatively) cold vapours of the chromosphere surrounding the hot body, or photosphere, of the sun. Sufficient matter to produce such absorption does not exist in the space between the earth and the sun.

Now we can experimentally determine the kinds of radiation emitted by the hot vapours of the various elementary substances. And if it is found that any of these radiations are absent from the spectrum of sunlight it is to be inferred that the vapours of these substances (provided the cause is not terrestrial) are present in the regions immediately surrounding the sun. In this way it is found that a great many substances existing on the earth's surface are present in the sun in the form of vapour. The lines A, B (Fig. 100) are due to oxygen, but have been shown to be caused by absorption in the earth's atmosphere. The lines C and F are due to hydrogen. The (double) line D is caused by sodium vapour. The (triple) line *b* is produced by the vapour of magnesium. Some hundreds of lines are caused by the presence of the vapour of iron.

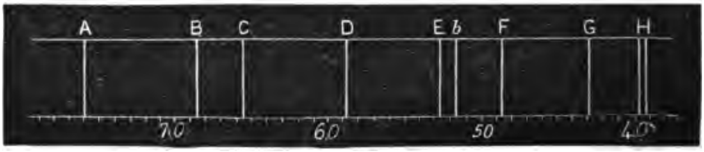


FIG. 100.

Approximate values of wave-lengths in centimetres may be got from the diagram by multiplying the scale numbers by $(10)^{-7}$.

In the same way the light emanating from any star, comet, or nebula, etc., may be examined, and the chemical constitution of the luminous body inferred. The various stars may be classified into several groups according to the nature of their spectra, and this classification indicates approximately their relative age. Generally speaking, the more recent stars have bright-line spectra, while the older stars exhibit continuous spectra crossed by numerous dark lines. The spectra of nearly extinct stars, however, resemble those of recent stars to a considerable extent.

If the slit of the spectrometer is wide the various coloured images overlap and produce an impure spectrum; and, under the same condition, a bright line broadens out and becomes indistinct. But, however narrow the slit may be, a line—even if due to light of one definite refrangibility alone—has always some finite breadth. The reason is that the radiating molecules are in violent motion—some moving towards, others moving from, the observer.

If n be the number of vibrations produced in the ether per second, and if the molecule emitting the light be at rest relatively to the observer, the wave-length, λ, of the disturbance is given by the equation $V = n\lambda$, where V is the speed of light. But, if the molecule be moving, relatively to the observer, with speed $\pm v$, the wave-length will be given by the equation $V \pm v = n\lambda'$; and the apparent change of wave-length is

$$\lambda' - \lambda = \pm v/n.$$

Hence, the molecules of a luminous body having all possible speeds included between the limits $+v$ and $-v$, a bright line in the spectrum of its light will possess finite breadth, even when it corresponds to one definite kind of radiation alone.

This principle has been applied to determine the rate of rotation of the sun on its axis, the speed of projection of gases in a solar eruption, and the rate of motion of stars to or from the earth.

Thus, Huggins has shown that Sirius is receding from the earth at the rate of about 35 kilometres per second.

Lockyer, by throwing light from the chromosphere on the slit of a spectroscope, the slit being placed radially with regard to the sun's disc, showed that some of the lines are broadened out in the lower parts of the chromosphere. This is an example of the broadening out of the lines of a gas under increasing pressure (§ 176). A similar broadening is observable in sun-spots, thus giving an independent proof of the fact that the sun-spots are depressions of the general surface.

The spectroscope is now a powerful aid in chemical research. It is said that it can indicate the presence of one three-thousand-millionth part of a gramme of sodium. A number of new elementary metals have been discovered by its aid.

179. *Law of Absorption: Body Colour. Dichroism.*—Let R be the amount of radiation of some definite kind which falls upon an absorbing medium. Let ρ (called the *absorption co-efficient*) be the percentage of this radiation, which is stopped by a plate of the medium of unit thickness. The quantity which passes through the given plate is therefore $R(1-\rho)$. A second plate of the substance, also of unit thickness, will stop the fraction, ρ of this quantity; so that the amount, $R(1-\rho)^2$, passes through a plate the thickness of which is two units. And, generally, the quantity which passes

through a plate, the thickness of which is n units, is $R(1-\rho)^n$. This practically vanishes, however small ρ may be (provided only that it is finite), when n is sufficiently great. Conversely, the amount of radiation, of the given kind, from a sufficient thickness of such a substance, is equal to that of a black body at the same temperature.

The expression $R(1-\rho)^n$ is only true on the assumption that ρ is constant for all radiations considered. If it is not so, we must take the sum of all such quantities instead.

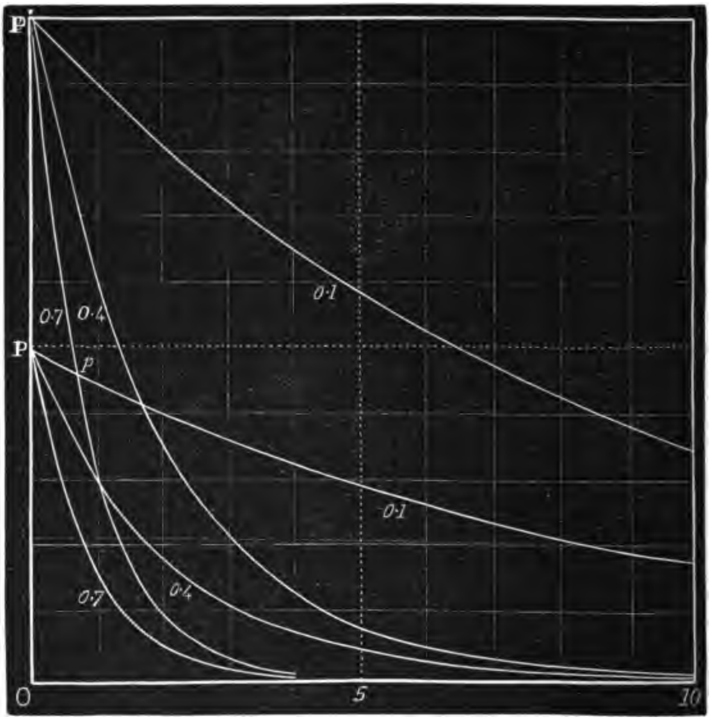


FIG. 101.

A substance which absorbs (say) red light, will appear bluish-green when viewed by transmitted light. And, the greater the thickness of the substance through which the light passes, the more monochromatic will be the apparent colour, until, finally, practically no light can pass.

Hence a substance into which light penetrates for a short distance and is then reflected out, will appear to be coloured, provided that selective absorption takes place, and its colour will be the same as that of the light which it transmits. This colour is termed the *body colour* of the substance.

For example, a mixture of blue and yellow pigments appears to be green because, if white light falls upon it, the particles of the blue pigment absorb the rays of small refrangibility, while the particles of the yellow pigment absorb the rays of large refrangibility. The green rays alone are partially reflected by both substances, and so the mixture appears to be green. On the other hand, a mixture of blue and yellow *lights* has a purplish tint. The only case in which the colour obtained by mixing two pigments is the same as that obtained by mixing the lights given out by the pigments individually is that in which the colour of the light absorbed by the one pigment is strictly complementary to the colour of the light absorbed by the other. But the purity of the mixed light is then inferior to that of the light reflected from the mixed pigments, for the white light which results from the combination of the complementary colours in the former case has its constituents entirely destroyed in the latter.

Now suppose that some substance absorbs (say) red light and green light, and let the coefficient of absorption for red light be much greater than the coefficient of absorption for green light. Suppose also that, in the incident light, the red rays are more intense than the green rays. It is obvious that, while the intensity of the red rays in the transmitted light will exceed the intensity of the green rays so long as the thickness of the substance is small, after a certain thickness is reached, the green light will be transmitted in greater intensity than the red light. The colour of such a substance, as seen by transmitted light, will therefore change from a reddish hue to a greenish hue as its thickness increases. This phenomenon is known as *dichroism*.

The accompanying diagram illustrates these facts graphically. The abscissæ of the curves represent the thicknesses of the absorbing medium; the ordinates of one set of curves represent the intensities of the transmitted light of one kind, and the ordinates of the other set indicate the intensities of the transmitted light of another kind, corresponding to the various thicknesses. The numbers accompanying the curves indicate different values of the coefficients of absorption. At the point p the high absorptive power (0·7) of the substance for the originally more intense light has diminished the intensity of that light to the same value as that which is exhibited by the originally feeble light, for which the coefficient of absorption is only 0·1.

Many such bodies occur in nature. For example, glass, coloured with a cobalt salt, while it transmits blue light when its thickness is small, appears red by transmitted light when its thickness is sufficiently great.

The law of absorption, above stated, *must* be true (neglecting such extraneous effects as internal reflection or scattering of light) so long as the coefficient of absorption does not depend upon the intensity of the light.

Sir G. Stokes has indicated a method of testing the point by the reflection of light, at nearly perpendicular incidence, from the surface of a sheet of glass, part of the absorbing medium being placed in the path of half of the light before reflection, and a similar part being placed in the path of the other (weakened) half after reflection. Both portions, being then projected on a screen, could be directly compared in respect of colour and intensity. Peddie used a method in which the light was passed through two double-image prisms (Chap. XVII.) and a plate of quartz. Four beams were thus produced, coloured alike in pairs, the colour of one pair being complementary to that of the other. The colour of one pair could be made to match as nearly as possible the colour most readily absorbed by the medium under test. The beams were made divergent by means of a lens which projected the light upon a screen. Similar portions of surface-coloured glass were placed in the paths of the two beams, which had the colour complementary to that of the glass, at different distances from the foci from which these beams diverged. In no case was any difference of effect produced, although the intensity varied from 1 to 1,000, and the eye could have perceived a difference of one hundredth part in the brightness of the two spots of light thrown on the screen, without the additional help of change of colour.

The meaning of the result is simply that each portion of the incident light produces its own proper effect independently of the simultaneous action of other portions. It is a result, therefore, which was to be expected, since all the known phenomena of light take place in accordance with the principle of superposition of effects.

180. *Surface Colour: Metallic Reflection.*—Some substances reflect from their surface certain rays only; thus gold reflects yellowish rays, and copper reflects reddish rays. The colour produced by this '*metallic reflection*' is called *surface colour.*

The light which is transmitted by a thin film of such substances is complementary to that which is reflected, that is, the transmitted light and the reflected light together make up a light of the same composition as that which was incident upon the surface. The reflected light cannot be plane-polarised at any angle of incidence (Chap. XVII.).

Many substances, besides metals, exhibit surface colour — for example, thin films of rose aniline, or of blue aniline, etc., appear of different colours according as they are viewed by transmitted, or by reflected, light. Such films may be prepared by placing a layer of an alcoholic solution of the aniline on a plate of glass and allowing the alcohol to evaporate. The colour, as seen by reflected light, varies somewhat with the angle of incidence.

The light reflected from such films cannot be entirely polarised at any angle of incidence. It consists of two parts—a part which can be plane-polarised at a certain angle of incidence and is identical with the transmitted light (which, in fact, constitutes the body colour of the substance)—and a part which cannot be plane-polarised, and so resembles the surface colour of metals. The polarisable part may be got rid of by suitable means, so that the remaining part may be examined alone. In the case of permanganate of potash, Stokes found that the surface colour seemed to be due to precisely those rays which were absent from the transmitted light, or, which is the same thing, the body colour. Hence, the colour of the light transmitted through this substance is due only to a very slight extent, if at all, to absorption. The spectrum of the transmitted light has five dark bands in the green part; the reflected light is green, and the spectrum of the surface colour portion of it consists of five bright bands, which correspond to the dark bands in the spectrum of the transmitted light.

181. *Anomalous Dispersion.* — In close association with the existence of dark absorption bands appears the phenomenon of *anomalous*, or *abnormal* dispersion.

In general, the rays of greater wave-length suffer refraction, on passage through a prism, to a smaller extent than the rays of shorter wave-length. But, in many substances, this rule does not hold. Such media are said to possess the property of anomalous dispersion.

Fox Talbot was the first to observe the phenomenon, but he did not publish his observations for about thirty years. In the meantime Le Roux had observed that iodine vapour refracted red light more than it refracted blue light.

Christiansen, Kundt, and others have widely extended our knowledge of such substances. Kundt has shown that the property of anomalous dispersion is possessed by all substances which exhibit surface colour.

If a continuous spectrum, such as may be given by a glass prism, be examined through another prism of a substance which exhibits abnormal dispersion, the spectrum will no longer be continuous but will present one or more dark bands. If the second prism be now turned so as to have its edge at right angles to the edge of the glass prism, the parts of the continuous spectrum will be displaced from their original positions to an extent depending upon the refractive index of the substance for each kind of light. The displacement of the rays, in a part of the spectrum close to a dark band, but of smaller wave-length than the absorbed rays, is abnormally small; and the displacement of the rays of slightly larger wave-length than those which are absorbed is abnormally great (Fig. 102).

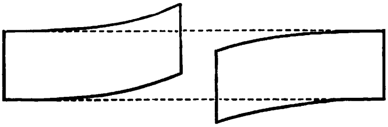


FIG. 102.

The general law, as given by Kundt, is that the rays of slightly less refrangibility than the absorbed rays have their refrangibility abnormally increased, while the rays of slightly greater refrangibility than the absorbed rays have their refrangibility abnormally diminished on passage through the absorbing medium.

182. *Fluorescence.*—The phenomenon of fluorescence is also necessarily associated with the absorption of light.

Brewster observed that the path of a beam of white light through a solution of chlorophyll glows with red light, and he termed the phenomenon 'internal dispersion.' Then Herschel noticed that the surface of a solution of sulphate of quinine upon which sunlight falls is of a bright blue colour. He named this appearance 'epipolic dispersion'; but Brewster showed that the blue colour could be manifested in the interior of the liquid if the beam of sunlight were sufficiently concentrated, and so he concluded that the phenomenon was of the same kind as that which he had already observed in the case of chlorophyll and fluorspar, etc.

Stokes has shown that a great many ordinary substances, such as bone, white paper, etc., possess this property, to which (avoiding any reference to dispersion) he gave the name of *fluorescence*, from its being noticeable in fluor-spar. His method of observation consisted in allowing a beam of light to enter a darkened chamber through a plate of blue cobalt-glass. This beam fell partly upon a white non-fluorescent body (white porcelain), and partly upon the body under examination. The light reflected from the two substances was then examined through a slit and prism. In this way the fluorescent light was compared with the light which produced it.

Stokes found that the light which was emitted by the fluorescent body was always of lower refrangibility than the light which produced it. The fluorescence of sulphate of quinine is due to the extreme violet rays of the spectrum, and to invisible rays of still higher refrangibility. Hence, by means of such a solution, the absence of rays beyond the visible part of a spectrum may be determined. For, if the spectrum be thrown on a screen damped with this solution, fluorescence is produced, beyond the usual visible part, except where rays of certain refrangibilities may be absent. In the case of chlorophyll the light which produces the effect is chiefly in the visible spectrum.

The explanation of the phenomenon given by Stokes is that the ethereal vibrations are absorbed by the fluorescent matter, which is set in vibration, the period of its vibration being usually longer than, never shorter than, the period of vibration of the ether. The vibrating matter now reacts upon the ether, and sets up in it vibrations which are generally longer still, but are never shorter than those induced in the molecules of the matter. This explains the lowering of refrangibility.

Dynamical illustrations of such interaction can be given. The following is due to Stokes. Ships at rest on a calm sea may be set in vibration by waves of definite period propagated from a distance. The natural period of oscillation of each ship will not generally agree with that of the waves. Any ship which is thus set in vibration will, by its vibrations, produce waves which spread outwards from it; but the period of these waves will generally be greater than that of the original waves, and can never be less than it.

If Stokes's explanation be true it is to be expected that the light which gives rise to fluorescence will be absent from the absorption spectrum of the substance. This is invariably the case.

Phosphorescence is precisely the same phenomenon as fluorescence. The only difference which subsists between the two is a difference of duration. Phosphorescence (so-called) frequently lasts for hours after the stimulating radiation is removed; fluorescence is maintained usually for only a small fraction of a second after the light ceases to fall on the substance.

Becquerel demonstrated and measured the finite time of duration of fluorescence in many substances. His apparatus consisted of a box with perforated revolving discs at either end. The perforations were so arranged that one end of the box was closed while the other was open. The substance which was to be tested was placed inside the box, and, on the discs (which had a common axis) being rotated, an intermittent beam of light passed through the substance. No light could pass out at the end of the box opposite to that at which the light entered unless the substance was fluorescent. But, that condition being satisfied, light could pass through when the speed of rotation of the discs was sufficiently great.

The duration of fluorescence is exemplified in the above dynamical illustration by the continued oscillation of the ships for some time after the cessation of the disturbance which originates it.

It may seem that the existence of fluorescence violates both Prevost's theory that the radiation from a body depends only upon its nature and its temperature, and the extension of that theory involved in the assertion of the equality of radiating and absorptive powers at a given temperature and for radiation of a definite wave-length. On this point Stewart remarked: 'I think we shall find, on examination, that in this general law it is taken for granted that no chemical change is taking place in the body in question, and no other molecular change than that implied in the cooling of the body. In a chemical action we have generally the transmutation of chemical energy into heat, and in molecular action we have generally the transmutation of molecular energy into heat likewise. That is to say, the body undergoing these changes becomes heated, and so gives out light and heat peculiar to the temperature to which it has been raised. But there seems to be no reason why molecular energy should not be somehow changed at once into radiant light and heat. In this case there would no doubt be an apparent breaking of the law above-mentioned, which associates a certain temperature with a certain quantity and quality of radiant heat, but the exception would be only apparent, for, as we have stated, the law presupposes that no molecular change of this nature is taking place.'

Perhaps the simplest way of stating the matter is to say that here we have virtually a violation of the condition that the enclosure must be totally impervious to radiation. For the introduction of a fluorescent body, within the enclosure, really enables some radiation to leave the enclosure in a potential form and other radiation to enter it from that form.

183. *Theories of Dispersion.*—Cauchy was the first to advance a

dynamical theory of dispersion. He ascribed it to the coarse-grainedness of the matter of which the dispersing substance is composed. The great difficulty of this theory is that, in order to account for the observed values of the refractive indices of substances such as glass, etc., the number of molecules of matter existing side by side in the length of a wave of light must be assumed to be much smaller than, from other considerations, can possibly be admitted. As already stated (§ 122), Lord Kelvin has recently shown that Cauchy's hypothesis can be so modified as to enable it to surmount this difficulty.

This hypothesis leads to an expression for the refractive index, μ, of any substance, of the form

$$\mu=a+\frac{b}{\lambda^2}+\frac{c}{\lambda^4}+\ldots\ldots$$

where a, b, and c, etc., are constants, and λ is the wave-length. This formula shows that the refractive index increases as the wave-length diminishes. Its results accord very well with experimental observations within the range of the visible spectrum, but it does not apply well to the invisible rays at the less refrangible end of the spectrum. The various terms rapidily diminish in numerical magnitude.

Briot generalised Cauchy's investigation somewhat, and deduced the expression

$$\frac{1}{\mu^2}=x\lambda^2+a+\frac{b}{\lambda^2}+\frac{c}{\lambda^4}+\ldots\ldots$$

which agrees better with experimental observations than the former does, and applies to a much greater range of wave-lengths.

The term $x\lambda^2$ depends upon the direct action assumed to exist between the ether and matter.

Modern theories (for example, that of v. Helmholtz) have regard, not so much to space relations—between wave-length and molecular distance—as to time relations—between the periods of vibrations in the ether and the periods of free oscillation of the material molecules.

V. Helmholtz assumes the existence of a viscous resistance to the motion of the molecules. When the periods of the ethereal and the molecular vibrations are identical, or approximately identical, absorption takes place, and, because of the viscosity, the vibrational energy takes the form of heat.

Kelvin's results differ from those of v. Helmholtz chiefly because he purposely avoids the assumption of the existence of viscosity. He obtains the equation

$$\mu^2=1+\frac{c_1\tau^2}{\rho}\left(-1+\frac{q_1\tau^2}{\tau^2-x_1{}^2}-\frac{q_2\tau^2}{x_2{}^2-\tau^2}-\ldots\ldots\right)$$

where μ is the refractive index, τ is the period of vibration of the ether, and x_1, x_2, etc., are the natural periods of oscillation of the molecules arranged in ascending order of magnitude. So long as τ is considerably greater than x_1 and considerably less than x_2, this equation will correspond to the case of ordinary refraction. As τ approaches x_1 in value the refractive index is abnormally increased. When τ is less than x_1, μ^2 is at first negative, but afterwards becomes positive, though abnormally small, as τ still further decreases. This explains the existence of anomalous dispersion. Negative values of μ^2, which accompany anomalous dispersion, indicate the existence of absorption or metallic reflection.

Thus the high reflecting power of silver is, on this theory, due to the fact that each one of all the kinds of radiation which are observed to be reflected from it has a vibrational period which is smaller than the smallest of the natural periods of oscillation of the molecules of silver.

Again, when τ has such a value that μ^2 is positive, but is less than unity (compare Kundt's observations, § 162), the particular radiation, of which τ is the period, will pass through the substance more quickly than it passes through air.

The energy of the rapid vibrations of the molecules is gradually transmuted into energy of the slow vibrations. This explains fluorescence and the radiation of heat from a body which has absorbed light. The molecule may be so constituted that the fluorescence (or phosphorescence) may last for a very long time.

CHAPTER XVI.

INTERFERENCE. DIFFRACTION.

184. *Principle of Interference.*—If light consists of undulations, propagated through the ether, the effects of which, at any point of the ether, are superposed in precisely the same way as are the effects of separate simple harmonic motions (§ 34), we should expect that conditions might occur under which the resultant motion at that point would be null—while, under other conditions, the resultant motion might be exceptionally great. We already know that, for a similar reason, when waves are propagated along the surface of water from two different sources, no resultant disturbance of the surface may exist at certain points. So also sounds from two different sources may be totally unheard by an ear placed at certain positions within hearing distance.

In order to produce continuous interference at given points it is absolutely necessary that the waves diverging from two sources should be of precisely the same period, as otherwise the resultant disturbance would vary from a minimum to a maximum alternately. Thus, in the case of sound (§ 150), difference of period gives rise to beats which may be observed by the ear.

Now the phase of the vibration emitted from one point of a flame has absolutely no relation with the phase of that emitted from any other point; and hence we cannot expect observable interference between rays coming from different luminous sources. Interference of course does occur between such rays constantly, but, in general, the alternations between maximum and minimum effects will succeed each other so rapidly that the eye can perceive no variation of intensity.

Therefore we conclude that, in order that persistent interference effects may be observable, the two interfering rays must originally have proceeded from a common source.

More that two centuries ago (1665) Grimaldi observed that, when rays of light from two sources overlapped each other and fell upon a screen, the portion of the screen which was illuminated by the two rays appeared to be darker than it was when illuminated by one ray alone. He allowed sunlight to enter a darkened chamber through two small apertures in the shutter. But these apertures were illuminated by light coming from all portions of the sun's disc, and so the effect which Grimaldi observed, to whatever cause it may have been due, could not have been produced by interference. Grimaldi, indeed, was not looking for interference phenomena—this was not thought of until 150 years later—he wished to prove that light was not material, since two portions of light apparently destroyed each other. And this reasoning is practically conclusive, for the conditions which would have to be assumed, in order to make an explanation of these phenomena by the emission theory possible, would be so arbitrary and artificial that no one could seriously advance them.

185. *Young's Experiment.*—Young, in 1801, was the first to observe true interference effects. He admitted light through a single small aperture in a shutter, behind which he placed another shutter pierced by two small openings. In this way he obtained two rays of light which proceeded originally from a common source—the single opening in the first shutter. That portion of a screen which was illuminated by both rays was crossed by alternately-arranged dark and bright bands. Young observed that the bands became narrower when the distance between the holes in the second screen was increased. He also noticed that the effect disappeared if either opening were closed.

The wave theory affords a ready explanation of the phenomena which Young observed.

Let A, A′ (Fig. 103) represent the two openings in the screen, and let AP, A′P be two rays which each illuminate the point P of the screen PN. M is the central point of AA′, and MN is drawn perpendicular to AA′ and PN. Denote the length of AM (or A′M) by a and the length of MN by b, and let x represent the distance PN; and let it be given that a is very small in comparison with b, while x is very small in comparison with a.

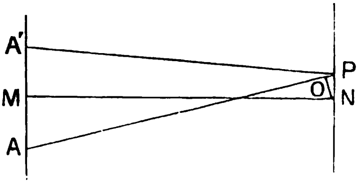


FIG. 103.

The waves, which travel along AP and A′P, start from A and A′ respectively in the same phase. Consequently the point P will be bright or dark according as AP − A′P is an even, or an odd, multiple of half a wave-length.

If we keep AP − A′P constant while the distance between the screens is varied, the locus of P is a hyperbola with A and A′ as foci. But, since the distance of P from M is very large in comparison with AA′, we see that, in the neighbourhood of P the hyperbola practically coincides with its asymptotes. That is, P practically lies on the straight line MP.

Now draw NO perpendicular to AP. Since P is very close to N in comparison with the distance a, and still more so in comparison with the distance b, OP is the excess of AP over AN or the defect of A′P from A′N, we get $2\text{OP} = n\lambda/2$. But the angle OPN is equal to the angle between AP and MN which is practically equal to ANM. Therefore $\text{OP}/\text{PN} = \text{AM}/\text{MN}$: whence we get

$$4ax = nb\lambda,$$

and the point P is bright or dark according as n is an even or an odd integer.

A and A′ may, of course, represent narrow luminous strips with their lengths perpendicular to the plane of the paper. The point P then corresponds to a dark or bright band also perpendicular to the plane of the paper.

By measuring the quantities a, b, and x, and by counting the number, n, of the particular band under observation, we can calculate the value of λ.

The distance between two consecutive bands is independent of n, and is therefore constant when a, b, and λ are fixed.

186. *Fresnel's Experiment.*—In Young's experiment the beams of light passed through apertures cut in a solid. Hence the observed effects might have been due to diffraction (§ 198). The result was that Young's explanation was not generally accepted; but a modification of his experiment, made by Fresnel, completely settled the matter.

Light, diverging from the point E (Fig. 104), is reflected from two mirrors, OR, OS, which are hinged together at O, and are inclined to each other at a very small angle. After reflection the rays appear to diverge from A and A′, the images of E in OS and OR respectively. Hence A and A′ act as two sources of light, the radiation emitted from each of which is similar in all respects to that emitted from the other. The light has nowhere passed through an aperture, so that the objection made to Young's form of the experiment does not apply, and yet the same effects are observed to occur.

The points A′, A, and E obviously lie on a circle, the centre of which is at O; and the lines OR and OS are respectively perpendicular to A′E and AE. Hence the angle A′EA is equal to the angle of inclination of the mirrors $=\theta$ (say). But $A'OA=2A'EA=2\theta$; and

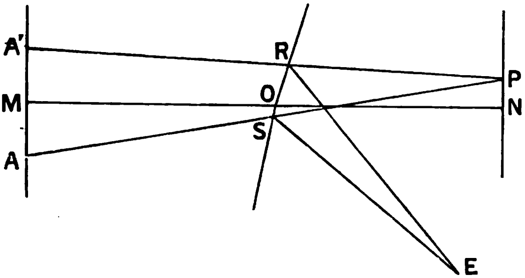


FIG. 104.

OM is practically equal to $OE=r$ (say). Therefore, if we denote ON by r', the formula of last section becomes

$$4r\theta x=n(r+r')\lambda$$

Very accurate adjustments are necessary in order to obtain good results from this form of the experiment.

187. *Lloyd's Experiment.*—Lloyd repeated the above experiment with only one mirror. A ray of light diverging from a slit, A′ (Fig. 105), is reflected in part, at grazing incidence, from a mirror

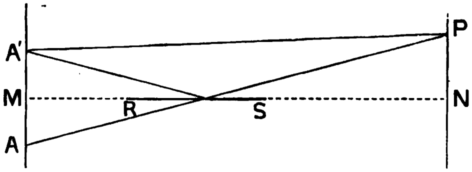


FIG. 105.

RS. We thus obtain two rays of light, one actually diverging from A′, and the other apparently diverging from A, the image of A′ in RS; and these rays produce interference effects as formerly.

Yet one distinct difference is observable. In both forms of the experiment previously described the point N is brightly illuminated, for AN − A′N = 0. In Lloyd's experiment N is dark, and the whole system of bright and dark bands is shifted by the breadth of one band. In explanation of this, Lloyd suggested that the phase is altered by 180° in the act of reflection.

In all cases the slit through which the light passes should be narrow; but this is not of so much importance in the present case as in the previous cases. For the slit A is the inverted image of A′, and so M is the centre of all corresponding parts of A and A′, the part of A which is nearest to M being the image of the part of A′ which is nearest to M, and so on. Hence the effects of all the parts are strictly superposed at P. But, in the two previous cases, since there is no inversion of A′ with respect to A, the part of A′ which is nearest to M corresponds to the part of A which is farthest from M (M being taken as the middle point of the line joining the central parts of the slits), and so on. Hence the systems of bands due to the light from the various corresponding parts of the two slits are not exactly superposed, and the definition is in consequence less accurate.

188. *Fresnel's Biprism.*—A second form of the experiment, to which also the objection taken to Young's experiment does not apply, is due to Fresnel. RS (Fig. 106) is a glass prism with a very obtuse angle. It is placed with its flat face towards M, the source of light. Each half of the prism forms an image of M, so that the rays emerge from the other faces of the prism as if they proceeded from points A and A′, which are practically situated on a straight line, through M, drawn perpendicularly to the flat face of the prism. If i_1, r_1, are the angles of incidence and refraction at the flat face of the prism, while i_2, r_2, are the similar angles at the opposite face, the total deviation of the ray MR, *i.e.*, the angle A′RM is (§ 167) $i_1 - r_1 + i_2 - r_2$. These angles being small, i_1 and i_2 are respectively equal to μr_1 and μr_2, μ being the refractive index of the substance of which the prism is composed. Hence the deviation is $(\mu - 1)(r_1 + r_2) = (\mu - 1)\alpha$, where α is the acute angle of the prism. This gives A′M(= AM) = $c(\mu - 1)\alpha$ approximately, c being the distance of

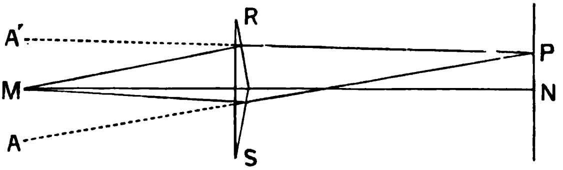


FIG. 106.

M from the biprism, and so the formula giving the value of x becomes

$$4c(\mu-1)ax=nb\lambda.$$

189. *Coloured Interference Bands: Michelson's Method.*—In the immediately preceding sections we have assumed the wave-length to be constant. But the breadth between two adjacent bright or dark bands is proportional to λ, and so the band situated at N is the only one which is colourless when white light is used. All other bands are coloured, the first red band being about twice as far from N as the first violet band. About a dozen of these bands can be fairly well distinguished when ordinary white light is used; but the succeeding bands of different colours are so superposed that all traces of interference effects practically disappear, and the screen seems uniformly illuminated.

If the quantity a in the formula of § 188 were variable and proportional to λ, x would be constant for all wave-lengths, that is, the bands would be colourless. This effect may be attained by the use of a diffraction grating (§ 207).

In the biprism method the distance between the points A and A depends upon the wave-length, and is greater the shorter the wave-length is. The result is that the coloured bands are more widely separated than they usually are when other methods are employed.

The introduction of a coloured glass, which diminishes the number of different kinds of light in the interfering beams, produces a very marked increase in the number of bands which are visible. Fizeau, using the light from sodium burning in air, succeeded in observing interference when the difference of path was as great as 50,000 wave-lengths. The number has been increased to fully 200,000 by the employment of sodium vapour under diminished pressure, since diminution of pressure makes the bright lines in the spectrum of a burning gas become sharper and narrower (§ 176), so that the light corresponding to each line becomes more completely monochromatic.

When the difference between the lengths of the paths travelled by the two interfering rays is a very large multiple of the wave-length, the nature of the vibrations may have completely altered in the interval of time between the setting out, from the source, of the two waves which simultaneously reach P, so that no interference could occur. But the fact that no more than 200,000 bands have ever been counted does not prove that no more than 200,000 vibrations of the ether at a given point are sufficiently nearly similar to produce continued interference, for we can neither obtain absolutely monochromatic light nor use an infinitely narrow slit. Yet the converse statement, that 200,000 successive vibrations are practically similar, is true.

In the case of the sodium light, which has two definite wave-lengths, the light of each wave-length gives rise to its own set of interference bands, and the resultant phenomenon is due to the superposition of these two sets. Since the difference of wave-length of the two components is very small, the first bright bands of the one set will be practically superposed upon the first bright bands of the other set; but gradually, as the total difference of path of the two interfering rays for each kind of light becomes greater and greater, the bright bands of one set will be superposed upon the dark bands of the other, and the screen will be uniformly illuminated. At greater distances from the central band there will be a gradual return to the first condition, and so on. The result is that there is a regular alternation of regions of strongly-marked bands with regions of more or less uniform illumination. If the source contains light of a number of different wave-lengths, the phenomenon will be correspondingly complicated; but it may still be possible to discriminate the different sets of bands in the same way as it is possible to discriminate the component simple harmonic disturbances in a complex wave. This method, which is due to Michelson, is by far the most delicate method of investigating the complexity of a source of light. Thus, for example, Michelson finds that, in the case of the red line of hydrogen, the first position of uniform illumination occurs where there is a difference of path of about 15,000 wave-lengths, and that the second similar position occurs where the difference of path is about 45,000 wave-lengths. This indicates that the red hydrogen line, which no spectroscope can resolve into components, is really a double line, the distance between the components being about one-sixtieth part of the distance between the components of the yellow sodium line.

In making such observations Michelson employs a special apparatus. A beam of light from the source passes, say, horizontally from west to east through a vertical piece of plane parallel glass which is inclined to the east and west line at an angle of forty-five degrees. This glass is silvered thinly on the side remote from the source, so that it reflects a considerable portion of the light in a south to north direction from its second surface. The light which is transmitted passes through a similar, and similarly placed, but unsilvered, piece of glass, and has its path exactly reversed by normal reflection from a metallic mirror. When this ray again reaches the silvered surface of the first piece of glass, it is reflected southwards to the observing telescope or the eye. The ray which was reflected northwards from the silvered surface is reversed by normal reflection at a second metallic mirror, and reaches the eye after again passing through the first glass plate. Each ray has passed through the same total thickness of glass; but the north and south ray was reflected, by internal reflection, at the surface of the first piece of glass, while the east and west ray was reflected, by external reflection, at its surface. Hence (§ 192) they interfere in opposite phases if the two totally reflecting mirrors are situated at equal distances from the point at which the reflection from the silvered glass plate occurs, and so the field is dark. If one of the mirrors be moved parallel to itself by a fine screw, two alternations of brightness and darkness will take place while the total path of the corresponding ray is altered by one wave-length. Thus, the screw may be calibrated in terms of wave-lengths by counting the number of alternations which take place in a given number of turns. Michelson and Morley have shown that by this method a given length may be estimated in wave-lengths, with an error of probably not more than one part in ten millions; so that we have a practical means of using the length of a definite ray of light as the standard of length.

190. *Displacement of Bands by Refracting Media.*—If a dense medium be placed in the path of one of the two interfering rays, the whole system of bands will be displaced towards that side of MN on which the medium is placed. For if t be the thickness of a medium of refractive index μ, which is traversed by the ray, the effect is the same as if the ray had traversed a thickness, μt, of air. Thus the effective length of the path of that ray is increased by the amount $(\mu - 1)\,t$.

Let L (Fig. 107) represent the medium interposed in the path of

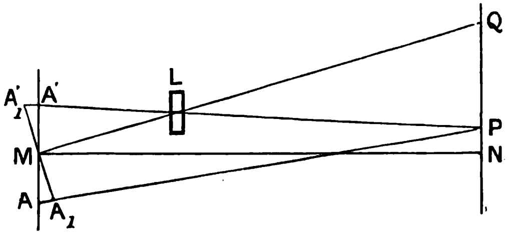


FIG. 107.

the ray A′P. The effective length of A′P is increased, and so the length of AP must be increased. In other words, PN must increase. Suppose now that L is removed, and that we shift A′ back from the screen through the distance $(\mu-1)t/2$ into the position A'_1. Let also A be moved towards the screen, through the same distance, into the position A_1. In this way the effective length of A'N is increased, relatively to that of AN, by the amount $(\mu-1)t$; and the central band, originally at N, will now be found at Q, which is such that MQ is perpendicular to $A_1A'_1$. But $QN/MN = AA_1/AM = (\mu-1)t/2a$. Hence the displacement of the central band (if we are dealing with monochromatic light) is

$$QN = \frac{b(\mu-1)t}{2a}.$$

When the light is not monochromatic the displacement of the central (which is then the brightest) band could only be given by this formula if the refracting substance did not produce dispersion, *i.e.*, if μ were independent of λ. The brightest effect will really be produced at a place where the rate of variation of QN with λ is a minimum, for, at such a place, the various adjacent coloured bands are most nearly superposed.

By means of the formula the refractive index of the interposed substance may be found with extreme accuracy. The method is specially applicable to the determination of the refractive indices of gases.

191. *Interference Bands in Spectra.*—If a pencil of rays of white light, diverging from a narrow slit, be made parallel by a suitable lens, and then be refracted by a prism, the usual continuous spectrum will be obtained. But if a plate of a refracting substance be interposed in the path of one half of the pencil, the spectrum will be crossed by dark bands. The reason is that one half of the rays are retarded relatively to the other half, and so interference effects are produced. Those rays, the relative retardation of which amounts to an odd multiple of a semi-wave-length, are obliterated.

Various forms of this experiment are described by Powell, Fox-Talbot, Brewster, and Stokes.

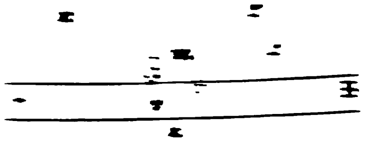
192. *Colours of Thin Plates. Reflected Light.*—Thin films of transparent substances are frequently observed to be brilliantly coloured. The colours vary with the angle of incidence and with the thickness of the film.

Familiar examples occur in the cases of a soap bubble, of the wing of the common house-fly, and of highly tempered steel, etc. In the latter case the thin film consists of an oxide formed on the surface of the steel at a high temperature. Very old glass vessels frequently exhibit these colours from the partial splitting away of thin films at the surface of the glass.

The wave theory gives a complete explanation of these phenomena.

[illegible] the [illegible] upper [illegible] is [illegible] in [illegible] the [illegible]

[illegible]



[illegible] feet [illegible] the [illegible] for [illegible] And [illegible] that [illegible]

[illegible] rays [illegible] integral [illegible] each [illegible] that. [illegible] semi-wave- [illegible] white [illegible]

[illegible] When the [illegible] and when [illegible] wave-length the [illegible] which [illegible] under faces. [illegible]

[illegible] place are [illegible] light is passing [illegible] passing from [illegible] analogy from [illegible] different manner. Young [illegible] which seems to be [illegible] of waves along a [illegible] is also precisely analogous. A wave propagated along the less dense portion is in part reflected from the junction with a complete reversal of phase. (As an extreme case imagine the rope to be fixed at the junction. This corresponds to infinite density of the second part.) A wave travelling along the more dense portion is partly reflected at the junction without change of phase.

Young pointed out that if his explanation were correct an entire reversal of the effects should occur when the reflecting plate was intermediate in density between the media on either side of it. Further, he carried out such an experiment, and found that his prediction was verified. Lloyd's experiment (§ 187) furnishes another verification of the correctness of Young's explanation.

The effective difference of path, $2\mu t \cos r$, decreases as the angle of incidence increases, and therefore the wave-length of the reflected light decreases as the angle of incidence increases. If the refractive index and the thickness of the plate be sufficiently large the series of colours may be repeated a number of times, but, if ordinary white light be used, partial overlapping will occur between all the series above the second, for the wave-length of the extreme red light of the spectrum is approximately double of that of the extreme violet light.

193. The above explanation of the reflection of light from thin plates is not quite complete. The intensity of the reflected ray, *bc*, is always greater than that of the ray *b′c′*, and so complete annulment of light is not accounted for. But complete annulment does take place. A complete treatment of the problem was given by Poisson, who pointed out that all the various rays which emerge at *b′* must be taken into account. The ray which enters at *b* and suffers one internal reflection at *d* before it passes out of the plate at *b′* has the greatest effect in producing the final result; but the ray which suffers three such internal reflections (at *e*, *b*, and *d*) before emergence also has a considerable effect. Similarly, those which have undergone five, seven, etc., internal reflections, have each an appreciable, though rapidly diminishing, share in the ultimate result.

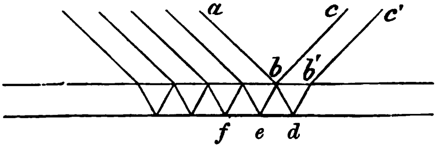


FIG. 109.

The effective difference of path between the ray which has suffered n such internal reflections and the ray which is once reflected externally at b is $(n+1)\mu t \cos r+\lambda/2$, the semi-wave-length being added in order to take account of the acceleration of phase produced in the act of reflection at b. This being taken into consideration, it is found that the intensity of the light reflected from the plate *does* vanish when $2\mu t \cos r$ is an even multiple of $\lambda/2$, and that it is a maximum when $2\mu t \cos r$ is an odd multiple of $\lambda/2$.

194. *Colours of Thin Plates. Transmitted Light.*—The light which is transmitted through the plate is complementary to that which is reflected from it; that is, the kinds of light which are absent from the reflected beam are precisely those which are present in the transmitted beam.

The intensity of the reflected light is never equal to that of the incident light, and so the intensity of the transmitted beam never entirely vanishes. Also, since the minimum intensity of the reflected light is zero, the maximum intensity of the transmitted light is equal to the intensity of the incident light.

195. *Newton's Rings.*—Newton observed the colours produced by the interference of rays reflected from both sides of a thin film of air enclosed between two pieces of glass. One of the pieces of glass had a plane surface; the surface of the other was convex and spherical.

The thickness of the film of air at a distance, d, from the point of contact is approximately $d^2/2R$ where R is the radius of the spherical surface. This follows at once from the fact that the product of the two segments into which an excentrical point divides the diameter of a circle is equal to the square of half the least chord through that point. Hence (see § 192) the condition that light of wave-length λ shall be intensified is

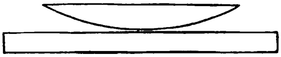


FIG. 110.

$$2\mu' d^2 \cos r=(2n+1)\lambda R,$$

n being an integer, and so the point of contact is surrounded by a series of bright rings. In this formula, r is the angle of refraction from glass into air and μ' is the reciprocal of the refractive index of glass.

It follows from this formula that the replacement of the film of air by a denser substance would cause all the rings to close in somewhat towards their common centre. This result is proved by experiment, and hence we get another proof of the fact that light travels slower through a medium such as water than it does through air.

The successive radii (that is, the successive values of d in the formula) are proportional to the square roots of the natural numbers —of the even numbers in the case of the dark rings, and of the odd numbers in the case of the bright rings, and so successive rings enclose equal areas.

The radii also increase as the wave-length increases, and so the first red ring is farther from the centre than the first blue one is.

Lastly, d increases when the angle of incidence increases, λ and n being constant.

When the two pieces of glass are pressed sufficiently close together a black spot appears at the centre. The central thickness is then very small in comparison with the wave-length of any visible light, and so the reflected light vanishes; for the effective length of the paths traversed by rays which emerge at a given point after internal reflection is practically the same as that of the light which is directly reflected at the same point without entering the thin film, and so the two sets of rays practically differ in phase by half a period.

The transmitted light is complementary to that which is reflected. The central portion is therefore white.

Theory indicates that if the refractive index of the film be intermediate between the indices of the two transparent media which bound it, the rings seen by reflection should commence from a white centre. Young verified this prediction by means of a film of oil of sassafras enclosed between a lens of crown glass and a lens of flint glass.

196. *Colours of Mixed Plates.*—If a bright object be viewed through an intimate mixture of two media of different refractive indices (*e.g.*, a mixture of oil and air enclosed between glass plates), colours are observed to which Young gave the name of 'colours of mixed plates.' The colours are arranged in rings precisely as in the case of those seen by transmission of light through a thin homogeneous plate, but the whole system is on a larger scale. The phenomenon is due to the interference of the rays which pass through the different media and so suffer relative change of phase.

When the incident light is oblique and a dark object is placed behind the plates, the system resembles that which is ordinarily seen by reflection, for one of the interfering portions is reflected and undergoes the usual acceleration of phase.

197. *Colours of Thick Plates.*—Brewster observed that, in certain circumstances, interference may be produced by means of plates, the

thickness of which is not small in comparison with the wave-length of light.

AB and BC (Fig. 111) represent two such plates of parallel glass, which are precisely equal in thickness, and are inclined to each other at a small angle.

A pencil of light, P*m*, falls perpendicularly upon the plate BC, and, passing through it, is partly reflected from the first surface of AB and in part is refracted into the plate AB. A portion of the refracted part is reflected at *m*. If r be the angle of refraction in AB the effective difference of path so produced between the two portions of light is $2\mu t \cos r$, where μ is the refractive index and t is the thickness of the glass. A similar action occurs at the plate BC, and the rays which were reflected from the first surface of AB sustain, relatively to the other rays, an effective increase of path to the amount $2\mu t \cos r'$, r' being the angle of refraction in BC. The effective difference of path of the two rays, *pq*, which finally emerge from the side of AB remote from P, is therefore the difference of these two quantities. Interference occurs when the difference is sufficiently small.

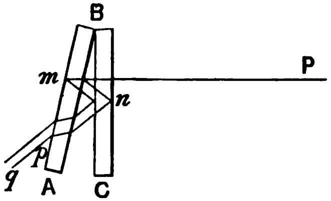


FIG. 111.

Jamin has applied this principle to the construction of an extremely sensitive instrument for the measurement of refractive indices.

Newton observed interference effects when he allowed light to fall upon the surface of a concave glass mirror which was silvered behind. The mirror was everywhere of uniform thickness and the light was admitted through a small opening in a sheet of white paper—the opening being situated at the centre of curvature of the mirror. A few broad coloured rings, resembling those due to light transmitted through a thin plate, were seen on the paper. All these rings were concentric with the opening through which the light passed.

The origin of these colours is totally different from that of the colours which Brewster observed. The rings are due to the interference of light, ordinarily reflected at the silvered surface of the mirror and then scattered by particles of dust on the first surface, with light, also reflected from the silvered surface, but which had been previously scattered (or, rather, *diffracted*) by particles of dust upon the first surface of the mirror.

When the mirror is slightly inclined the centre of the coloured rings is situated midway between the opening in the paper and the image of it which is formed upon the paper. This central spot is alternately bright and dark (where homogeneous light is used), as the distance between the opening and its image increases; it undergoes a rapid variation of colour when the incident light is white.

198. *Diffraction.*—The principle by means of which Huyghens explained the rectilinear propagation of light has already been given (§ 161). The following remarks of Stokes on this subject are specially worthy of note:

'When light is incident on a small aperture in a screen, the illumination at any point in front of the screen is determined, on the undulatory theory, in the following manner. The incident waves are conceived to be broken up on arriving at the aperture; each element of the aperture is considered as the centre of an elementary disturbance, which diverges spherically in all directions, with an intensity which does not vary rapidly from one direction to another in the neighbourhood of the normal to the primary wave, and the disturbance at any point is found by taking the aggregate of the disturbances due to all the secondary waves, the phase of vibration of each being retarded by a quantity corresponding to the distance from its centre to the point where the disturbance is sought. The square of the coefficient of vibration is then taken as a measure of the intensity of illumination. Let us consider for a moment the hypothesis on which this process rests. In the first place it is no hypothesis that we may conceive the waves broken up on arriving at the aperture: it is a necessary consequence of the dynamical principle of the superposition of small motions, and if this principle be inapplicable to light, the undulatory theory is upset from its very foundations. The mathematical resolution of a wave, or any portion of a wave, into elementary disturbances must not be confounded with a physical breaking up of the wave, with which it has no more to do than the divisions of a rod of variable density into differential elements, for the purpose of finding its centre of gravity, has to do with breaking the rod in pieces. It *is* a hypothesis that we may find the disturbance in front of the aperture by merely taking the aggregate of the disturbances due to all the secondary waves, each secondary wave proceeding as if the screen were away; in other words, that the effect of the screen is *merely to stop* a certain portion of the incident light. This hypothesis, exceedingly probable, *à priori*, when we are only concerned with points at no great distance from the normal to the primary wave, is confirmed by experiment, which shows that the same appearances are presented, with a given aperture, whatever be the nature of the screen in which the aperture is pierced; whether, for example, it consist of paper or foil, whether a small aperture be divided by a hair or by a wire of equal thickness. It is a hypothesis, again, that the intensity in a secondary wave is nearly constant, at a given distance from the centre, in different directions very near the normal to the primary wave; but it seems to me almost impossible to conceive a mechanical theory which would not lead to this result. It is evident that the difference of phase of the various secondary waves which agitate a given point must be determined by the difference of their radii, and if it should afterwards be found necessary to add a constant to all the phases the results will not be at all affected. Lastly, good reasons may be assigned why the intensity should be measured by the square of the coefficient of vibration.'

199. Huyghens' construction, if rigorously carried out, wou indicate the existence of a wave running back towards the source as well as a wave which travels forwards. Analogy points to the conclusion that the part of the construction which leads to a reverse wave must be ignored. For example, the investigation of § 56 shows that no wave can travel backwards from a disturbance which runs along a stretched cord. But Stokes, in his paper on the Dynamical Theory of Diffraction, of the introduction to which the above quotation forms part, has shown from purely dynamical principles, that the disturbance in a secondary wavelet is a maximum in the direction of the wave-normal, and that it diminishes constantly as the direction considered is inclined more and more to the normal, ultimately becoming zero in the direction opposite to that in which the primary wave travels. He then shows that the result of the superposition of all the secondary effects is the same as if the wave (assumed to be practically plane, *i.e.*, having a radius which is large in comparison with the wave-length, a condition always satisfied in experiments) had not been supposed to be broken up into a series of separate centres of disturbance, and that no back-wave is propagated.

200. *Effect of a Rectilinear Wave.*—We have already stated that Fresnel showed that Huyghens' principle, according to which the

new wave-front is found to be the envelope of the secondary wave-fronts, should be explicitly associated with the principle of interference if it is to give a complete explanation of the rectilinear propagation of light. The envelope is the locus of points each of which is simultaneously reached by more than one secondary disturbance the phases of which are identical. It is, therefore, the locus of points at which the light has great intensity.

The necessity for the introduction of the principle of interference will appear very evidently from the following investigation of the effect of a rectilinear wave at any external point.

Let AB (Fig. 112) represent a portion of a linear wave which extends to infinity in both directions, and let P be the point at which we have to determine the effect of the wave. Draw PM perpendicular to AB and take points m, m', m'', etc., such that $Pm - PM = Pm' - Pm = Pm'' - Pm' =$ etc., $= \lambda/2$, where λ is the wave-length of the light emitted from the various points of AB.

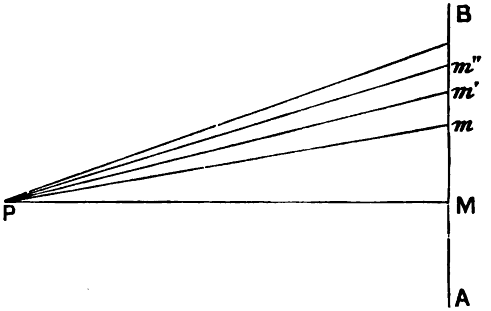


FIG. 112.

The length of the half-period element Mm is $\sqrt{(a+\lambda/2)^2 - a^2}$ where a is the length of PM. When λ is so small in comparison with the other length involved that it may be neglected, this becomes $\sqrt{a\lambda}$. Similarly Mm', Mm'', etc., are respectively equal to $\sqrt{2a\lambda}$, $\sqrt{3a\lambda}$, etc. Hence the lengths of the successive half-period elements, from M outwards, are $\sqrt{a\lambda}$, $\sqrt{a\lambda}(\sqrt{2}-1)$, $\sqrt{a\lambda}(\sqrt{3}-\sqrt{2})$, etc., and the limit to which they ultimately approach, as the line drawn from P to the element becomes more and more nearly parallel to MB, is $\lambda/2$.

If we divide each element into the same number of infinitesimal portions, the light sent out by the first portion of the first element differs in phase from that emitted by the first portion of the second element by one half of a period. Similarly, the light emitted by the second portion of the first element differs in phase by one half of a period from that emitted by the second portion of the second element, and so on. Now the effects at P of the various parts of the first element are not quite compensated by the effects of the corresponding parts of the second element. For the breadth of the parts of the first element is rather greater than that of the parts of the second; also the inclination of each part to the line joining it to P and the distance of each part from P increase as the part is more remote from M. But the difference between the effects of the corresponding portions of the n^{th} and the $n+1^{th}$ elements is vanishingly small when n is large.

Now, as λ is a very small length, it follows that a very large number of half-period elements is included in a small portion of AB in the near neighbourhood of M, and, consequently, only a small part of the wave near M produces any effect at P. Hence, a small opaque object placed on the line PM would entirely prevent the wave AB from producing any effect at P.

Hence the propagation of light is practically rectilinear when λ is so small that its square may be neglected in comparison with the other quantities involved.

Let e_1, e_2, etc., be the effects at P of the first, second, etc., half-period elements of MB. The total effect (taking account also of the portion MA) is

$$2(e_1-e_2+e_3-e_4+\ldots\ldots-e_{2n}+\ldots\ldots).$$

These various terms are in descending order of magnitude, and therefore it appears that the total effect is smaller than the effect of the first half-period portions at M.

The difference between any two successive terms is small in comparison with the magnitude of either, and so, writing the above expression in the form

$$e_1+(e_1-e_2)-(e_2-e_3)+(e_3-e_4)+\ldots\ldots,$$

we see that the total effect at P is approximately equal to the effect of one half-period element at M.

201. *Effects of Plane and Spherical Waves.*—Let the plane of the paper represent a plane wave, the effect of which at a point, P, is to be found, and let M (Fig. 113) be the foot of the perpendicular drawn from P to the plane.

From P as centre describe successive spheres of radii $MP+\lambda/_2$ $MP+2\lambda/2$, etc. The spheres will divide the plane wave into concentric zones, called half-period zones, or, sometimes, Huyghens' zones.

Dividing each of these zones into the same number of infinitesi mal annular portions, we observe that the effect of each portion of one zone is nearly annulled by the effect of the corresponding portion of the succeeding zone, and that the annulment is practically complete at a short distance from the point M—precisely as in the similar investigation of last section. Hence the effect produced at P is that due to a few half-period zones in the neighbourhood of the wave-normal which passes through P, and is practically equal to half the effect of the first zone.

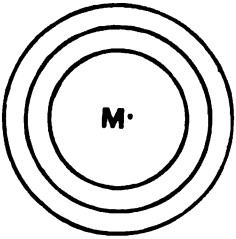


FIG. 113.

Let AMB (Fig. 114) represent a spherical wave diverging from O. To find the effect at P we must, as above, divide AMB into half-period zones surrounding M the point in which OP intersects AB.

Reasoning similar to the foregoing shows that the effect of the wave at P is equivalent to half of that produced by the first zone.

Now it is easy to see that the phase of the vibration due to the first zone differs from that due to the secondary wave at M by one quarter of a period. For if OM be large in comparison with the wave-length—a condition which is satisfied in all experimental observations—the first zone is practically plane. And, further, if it be broken up into $2n$ infinitesimal rings, of equal area, surrounding the point M as centre, the amplitudes of the vibrations produced

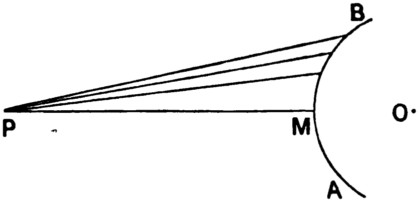


FIG. 114.

at P by each of these annular portions will be practically equal to one another. Hence (§ 52) the phase of the resultant of the effects of the 1^{st} and the $2n^{th}$ annuli is halfway between those of its components. This is true also of the phase of the resultant of the effects of the 2^{nd} and the $(2n-1)^{th}$ annuli, and so on. But the phase of the vibration at P due to each annulus varies uniformly from the 1^{st} to the $2n^{th}$ annulus. Therefore the phase of the resultant vibration at P due to the complete zone is one-quarter of a period behind that due to the vibration at M. The same statement must be true of the vibration produced by the whole wave if it agrees in phase with that produced by the first zone.

202. *Diffraction at a Straight Edge.*—We are now in a position to determine the effects produced by any given portions of a wave which diverges from a luminous point—the remaining portions being intercepted by opaque obstacles. This involves the carrying of our investigations beyond the stage in which the wave-length may be assumed to be small in comparison with all other quantities involved. We shall find that, under this new condition, light is no longer propagated in straight lines, but is bent, or *diffracted*, into the geometrical shadows precisely as sound is.

AMB (Fig. 115) represents a spherical wave, which, diverging from the point O, is partially intercepted by an opaque object MN.

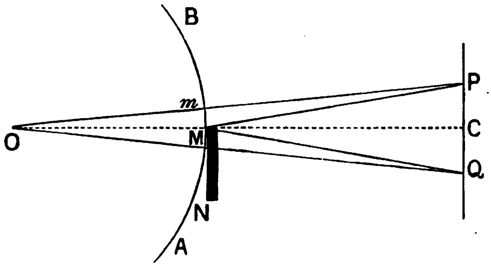


FIG. 115.

We have to determine (1) the effect at any point P outside the geometrical boundary, OMC, of the shadow; (2) the effect at any point inside Q the geometrical shadow.

Join OP and MP, and let OP meet AB in m.

When mM contains a considerable number of half-period elements, the wave produces practically its full effect at P. Let e_1, e_2, etc., be the effects of the first, second, etc., half-period elements in the neighbourhood of m; and let E be the effect of the semi-wave mB. The effect at P is E, E$+e_1$, E$+e_1-e_2$, etc., according as the number of elements included in mM is 0, 1, 2, etc. Hence the effect at P is a maximum, or a minimum, according as mM contains an odd, or an even, number of half-period elements; that is, according as, in the formula

$$\mathrm{MP}-m\mathrm{P}=n\frac{\lambda}{2},$$

n is odd or even. When n and λ are given, the locus of P is a hyperbola the foci of which are O and M. Now, if we denote PC_2 by x, OM by a, and MC by b, we get $\mathrm{OP}^2=(a+b)^2+x^2$ and $\mathrm{MP}^2=b^2+x^2$. These expressions give approximately $\mathrm{OP}=a+b+x^2/2(a+b)$, and $\mathrm{MP}=b+x^2/2b$. Hence the above formula becomes

$$x^2\frac{a}{2b(a+b)}=n\frac{\lambda}{2}.$$

At a point Q, within the geometrical shadow, the most effective portions of the wave are intercepted by the obstacle. The effect at Q is practically $\frac{1}{2}e_2$, $\frac{1}{2}e_3$, etc., according as MN intercepts 1, 2, etc., of the most powerful elements. Hence the illumination inside the geometrical shadow dies away as the distance of the illuminated point from the geometrical boundary increases.

Diffraction fringes, resembling those just described, appear outside the geometrical shadow on both sides of a narrow obstacle, such as a thin wire or a hair. But, in addition, a series of fine bands, of constant breadth, make their appearance inside the geometrical shadow if the obstacle is sufficiently narrow. These are caused by interference of the light diffracted at both sides of the obstacle, for, as we have seen, the effect of each unintercepted portion of the

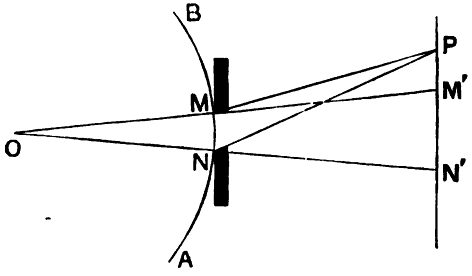


FIG. 116.

wave is practically the same as that of a luminous line placed close to the straight edge of the obstacle.

203. *Diffraction at a Narrow Slit.*—Let MN (Fig. 116) represent a narrow opening in an opaque obstacle, and let a wave AB diverge from a point O, which is situated on the line drawn from the middle point of MN at right angles to the plane of the obstacle.

Reasoning similar to that of last section shows that the illumination at a point P will be a maximum, or a minimum, according as MN contains an odd, or an even, number of half-period elements.

Let M′N′ be the geometrical projection of MN. If the screen, PM′N′, be so far from MN that NM′−MM′ (or MN′−NN′) is less than a semi-wave-length, a fringe of alternately bright and dark bands will appear on each side of the geometrical projection of MN. But if the distance between the obstacle and the screen be so small that NM′−MM′ is greater than a semi-wave-length, bands will appear between M′ and N′.

204. *Diffraction at a Circular Aperture. Zone Plates.*—Draw the line OP (Fig. 117) through the centre of the aperture and perpendicular to its plane. We shall determine the general effect at P of light diverging from O.

The illumination at P is a maximum, or a minimum, according as MN contains an odd, or an even, number of half-period zones.

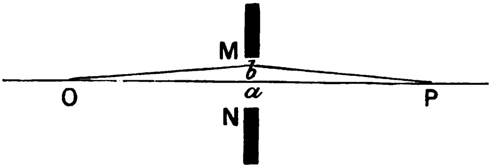


FIG. 117.

Let a be the centre of the aperture, and let b be the outer edge of the n^{th} zone. Denote the lengths of Oa, aP, and ab by u, v, and x respectively. We get approximately $Ob = u + x^2/2u$ and $bP = v + x^2/2v$; therefore

$$Ob + bP = u + v + \frac{x^2}{2u} + \frac{x^2}{2v}.$$

Hence

$$Ob + bP - OP = \frac{x^2}{2}\left(\frac{1}{u} + \frac{1}{v}\right).$$

But this length is equal to $n\lambda/2$, and so

$$x^2\left(\frac{1}{u} + \frac{1}{v}\right) = n\lambda.$$

Thus the consecutive values of x are proportional to the square roots of the natural numbers.

As P approaches MN, the number of half-period zones in the aperture increase, and so the illumination at P passes through a succession of maxima and minima. The last equation shows that the various points at which the maxima and minima occur are given by the expression

$$v = \frac{ur^2}{un\lambda - r^2}$$

where r is the radius of the aperture.

From the way in which λ is involved, it is evident that the position of P corresponding to maximum illumination approaches nearer to the aperture when the wave-length increases; so that an eye, which advances to the aperture along Pa, will perceive a rapid periodic variation in the colour of the light which reaches it.

If, as formerly (§ 200), we denote by e_1, e_2, etc., the effects of the light which passes through the successive half-period zones, the total effect is

$$e_1 - e_2 + e_3 - e_4 + \ldots\ldots$$

The effect will therefore be much greater than it otherwise could be if the even zones be made opaque. Such an arrangement constitutes a *zone plate*. If n be the number of zones (alternately open and opaque) in a zone plate of radius r, the formula

$$\frac{1}{u} + \frac{1}{v} = \frac{n\lambda}{r^2}$$

shows (§ 168) that the plate acts as a condensing lens, the principal focal length of which is $r^2/n\lambda$. But, in the case of the zone plate, all rays do not take the same time to pass between the conjugate foci; and, further, the focus for red rays is nearer to the plate than the focus for blue rays is. In these points it differs from a lens.

205. *Diffraction at an Opaque Disc.*—A point at the centre of the geometrical shadow is almost as brightly illuminated as if the disc were removed. If this disc removes $n-1$ half-period zones, the effect of the remaining zones is practically equal to one-half of that of the n^{th} zone. But, so long as n is not large, the effect of the n^{th} disc is not greatly different from that of the first. This theoretical result was first pointed out by Poisson, and was verified experimentally by Arago.

206. *Coronæ. Young's Eriometer.*—If a number of very small and nearly equal particles be closely distributed in the space intervening between a luminous object and the eye, the object will appear

to be surrounded by luminous rings. These are due to diffraction of the light which passes the edges of the particles. The *coronæ*, which are sometimes seen surrounding the sun or the moon, are caused by the presence of small globules of water in the atmosphere. They are coloured blue inside, red outside, and increase in size when the diameters of the globules diminish. [If, therefore, the coronæ are observed to contract, the moisture in the atmosphere is condensing, and rain may be expected to follow ; conversely, if the rings dilate, dry weather will in general ensue.]

Young's *Eriometer* was devised for the purpose of measuring the diameters of small objects. It consists of a metal plate, in which a small hole is drilled. The plate is also perforated by a circle of still smaller holes which surround the large hole as a centre. A flame is placed behind this plate, and the light which passes through the holes is examined through glass plates which contain between them the (equal) particles the size of which is to be determined. The large opening in the metal plate is surrounded by coloured rings, and the distance between the metal plate and the glass plates is altered until some particular ring coincides with the circle of small holes. This distance varies inversely as the radius of the ring, which itself, as we have just seen, varies inversely as the diameter of the particles. One experiment, in which the diameter of the particles is known, and the distance between the plates is measured, is sufficient to enable us to calculate the unknown diameters of other sets of particles.

207. *Diffraction Gratings.*—A diffraction grating may consist of a glass plate, upon which a great number of extremely fine equidistant parallel lines are ruled by means of a diamond point. The grooves are practically opaque, for light incident on them is reflected back in all directions. On the other hand, the glass between the grooves is transparent, and the light which passes through is diffracted in all directions.

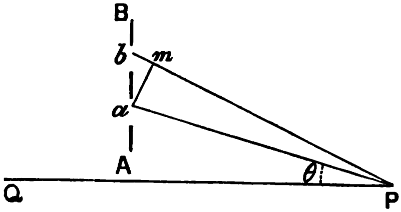


FIG. 118.

Let AB (Fig. 118) represent a (highly-magnified) portion of the grating, the dark spots indicating the grooves, and the light parts indicating the intervening spaces; and let aP and bP represent the paths of rays which reach P from similarly-situated parts of the openings a and b. From a drop a perpendicular am upon bP. The distance ab being very small in comparison with the distance of P from the grating, bm is practically equal to $b\mathrm{P}-a\mathrm{P}$.

Now suppose that parallel rays from a narrow slit parallel to the grooves fall perpendicularly upon the grating. An eye placed at P will see the slit through the grating in the direction PQ. The angle bam is equal to the angle $a\mathrm{PQ}(=\theta$, say); and so $bm=ab\sin\theta$. Therefore a maximum or minimun effect will be produced at P according as n is even or odd in the expression

$$ab\sin\theta=n\lambda/2.$$

The length ab is known, since it is the reciprocal of the number of grooves ruled in unit breadth of the grating.

If monochromatic light be used, a series of coloured images of the slit will be seen at different angular distances from the line PQ. If white light be used, a series of spectra will be observed, in each of which the violet light is less bent from its original direction than the red light is. The spectra are said to be of the first, second, etc., order, according as n has the values 1, 2, etc. All the spectra beyond the second partially overlap each other.

Very accurate measurements of wave-lengths may be made by means of the grating; and the spectra obtained from all gratings are identical, except as regards *scale*; that is, there is no trace of irrationality in the dispersion (§ 172). And, further, if θ is nearly zero, the dispersion between any two rays is practically proportional to the difference of their wave-lengths. This condition may be attained by inclining the grating to the direction of the incident light at a suitable angle. The spectrum thus produced is called a *normal spectrum.*

Diffraction spectra may be obtained by reflection from a ruled metallic surface. Rowland's concave gratings are ruled on the polished surface of a portion of a sphere of speculum metal. The earliest grating was made by Fraunhofer. He wound fine wire round two screws whose axes were parallel and which had 40 threads per centimetre of their lengths. Some of Rowland's gratings have more than 10,000 lines per centimetre.

CHAPTER XVII.

DOUBLE REFRACTION. POLARISATION.

208. *Double Refraction.*—In our consideration of the refraction of light, we have hitherto dealt only with those cases in which a single refracted ray occurs.

Bartholinus, in 1669, described the phenomenon of double refraction as observed by him in Iceland spar.

A single ray of light incident upon the surface of Iceland spar in general gives rise to two refracted rays. One of these obeys the ordinary law of refraction, but the other follows a totally different law. The former is called the ordinary, and the latter the extraordinary, ray.

All crystalline minerals, except those belonging to the cubic system, possess the property of double refraction.

The fundamental form in which Iceland spar crystallizes is the rhombohedron. The angles of the faces are either acute or obtuse. The obtuse angles are all equal to each other, and the acute angles are all equal also. Two of the solid angles (A and B, Fig. 119) of the rhombohedron are bounded by three obtuse angles. All other angles, such as C, are bounded by one obtuse and two acute angles. The *axis of the crystal* is a line which is equally inclined to the three edges meeting at an obtuse-angled corner. If we make all the edges of the block equal in length, the crystalline axis will be AB,

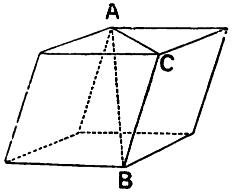


FIG. 119.

the diagonal joining the two obtuse-angled corners. A plane ACB, which passes through the crystalline axis and the shorter diagonal AC of a rhombic face, is called a *principal section* of the crystal.

In Iceland spar, and in many other crystalline substances, all the optical properties are symmetrical about the axis of form. Any *direction* in such a substance, which is parallel to the axis of form, is, therefore, called the *optic axis ;* and all such substances are called *uniaxal* crystals.

If the spar be cut by a plane in any direction, and a ray of light falls upon the surface so formed, both an ordinary and an extraordinary ray will in general be produced; and, in most cases, the latter will not lie in the plane of incidence. But if the plane be perpendicular to the optic axis, both rays coincide if the incidence is normal. This also occurs if the optic axis lies in the refracting surface, and the incidence is normal; and, further, in this case the extraordinary ray obeys the ordinary law so long as the plane of incidence is perpendicular to the optic axis.

These various phenomena were investigated very fully by Huyghens, and he was led to adopt a construction for the wave-front in the interior of the crystal which he himself proved experimentally to accord very accurately with the observed facts. More severe tests of his construction were made by Wollaston in 1802; and recently Stokes, Mascart, and Glazebrook have verified its accuracy to the full extent attainable by modern methods of measurement.

209. *Huyghens' Construction.*—Huyghens has previously explained the propagation of light in homogeneous isotropic media by the assumption that the wave-surface was spherical (§§ 161, 175). To explain double refraction in uniaxal crystals, he assumed that the wave-surface consists of an ellipsoid of revolution the axis of symmetry of which is coincident with the optic axis, and a sphere which touches the ellipsoid at the extremity of its axis of symmetry.

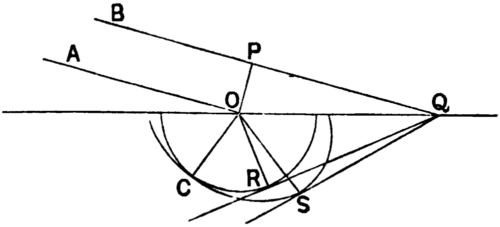


FIG. 120.

The spherical portion of the surface corresponds to uniform speed of propagation in all directions; and, the incident ray being given, we can determine from it, by the method of § 175, the direction of the ordinarily refracted ray. The ellipsoid indicates unequal speed of propagation in different directions, and from it we can determine the direction of the extraordinarily refracted ray by a similar process.

Let O (Fig. 120) be the point at which an incident ray AO meets the surface OQ, and let BPQ be another ray parallel to AO, so that OP, which is perpendicular to both, may represent a portion of a plane wave-front. In the time in which light moves from P to Q, the ordinary ray will have passed over a distance OR, such that PQ = μOR, where μ is the ordinary index of refraction. A plane through Q, perpendicular to the place of incidence, will touch a sphere drawn from O with radius OR in a point R, and OR is the direction of the ordinary ray. The plane RQ is the ordinarily refracted wave-front.

If OC is the optic axis, the radii of an ellipsoid CS, which has OC as its semi-diameter of revolution, will represent the speeds of propagation of the extraordinary ray in different directions. A plane passing through Q, and perpendicular to the plane of incidence, will touch CS in a point S such that OS is the direction of the extraordinary ray; and the ratio PQ/OS is equal to μ', the index of refraction for all extraordinary rays which pass through the crystal in the direction OS.

In Iceland spar, OC is the shortest radius of the ellipsoid; in quartz it is the largest radius. All crystals which resemble Iceland spar in this respect are called *negative* crystals; those which resemble quartz are called *positive* crystals. In the former, the extraordinary index is less than the ordinary; in the latter, the reverse is the case.

If, in this figure, the point C lies out of the plane of the paper, the point S will in general lie outside it also; that is, the extraordinary ray will not be in the plane of incidence. This will be so even if the incidence is perpendicular; for the new wave-front will be a plane parallel to OQ, and this will in general touch CS in a point which does not lie in the plane of the paper.

210. *Special Sections of the Surface.*—(1) Let the refracting surface be perpendicular to the optic axis (Fig. 121). At normal incidence there is no separation of the two rays; but, as the angle of incidence increases, the extraordinary ray separates out farther from the normal than the ordinary one does. If the plane of incidence be rotated around OC, the two rays each maintain a

fixed inclination so long as the angle of incidence remains constant.

(2) Let the refracting surface and the plane of incidence intersect

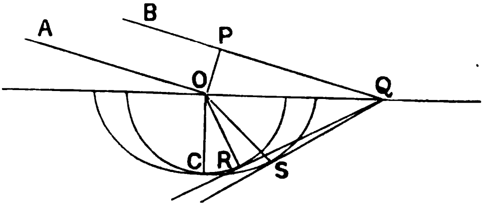


FIG. 121.

in the optic axis (Fig. 122). At normal incidence there will be no separation of the two rays as regards direction, though the extraordinary ray will travel with greater speed than will the ordinary

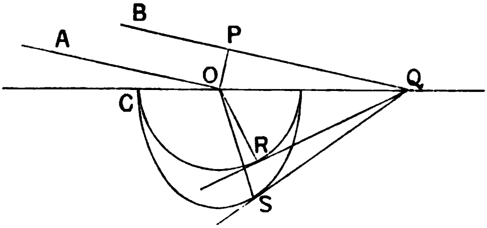


FIG. 122.

ray. And, when the angle of incidence increases, the former does not separate out so far from the normal as the latter does; for, from the properties of the ellipse and circle with a common diameter, R and S lie on a line which is perpendicular to OQ.

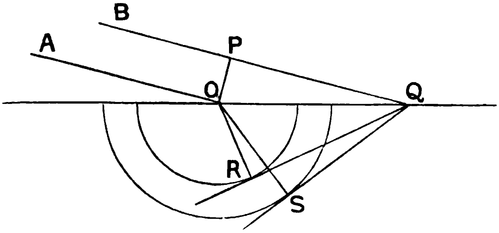


FIG. 123.

(3) Let the refracting surface contain the optic axis, while the plane of incidence is perpendicular to it (Fig. 123). The section of the

ellipsoid becomes a circle, and so the extraordinary ray obeys the ordinary law, though its index of refraction is less than that of the ordinary ray. At normal incidence, this case becomes identical with the last.

211. *Polarisation.*—Huyghens observed that the intensities of the two beams produced by refraction in a block of Iceland spar are equal. And he further noticed that each of these beams was in general subdivided into two others, of *unequal* intensity, on transmission through a second block.

When the principal sections of the two blocks are parallel, no more than two beams are produced: the ordinary ray in the first block passes through the second, without change of direction, as an ordinary ray; and the extraordinary ray passes through also without any change. And, when the principal sections of the blocks are at right angles to each other, two rays only are transmitted; but the ordinary ray in the first block passes through the second as an extraordinary ray, while the extraordinary ray in the first becomes an ordinary ray in the second. In all other relative positions of the two principal sections, each ray in the first is subdivided into two in the second. As the second block is turned round from the position in which its principal section was parallel to that of the first, the two original beams gradually diminish in intensity as the intensities of the newly-produced beams increase. When the principal sections are inclined at an angle of 45° to each other, all the four rays are equally intense. The changes then proceed in the same direction until the inclination of the principal sections is 90°, when the original beams vanish; after this, if the inclination be still further increased, the changes proceed in the reverse order until, at 180°, the beams again pass unchanged through the second block.

Huyghens remarked that the rays which had passed through the first block seemed to have acquired some *form or disposition* which led to the production of this phenomenon. Newton spoke of them as possessing *sides*. But it was not until more than a century afterwards that a complete explanation was found as the result of an accidental discovery.

Malus, happening to examine through a doubly refracting prism the light reflected from the windows of the Luxembourg Palace, observed that each ray alternately disappeared as he rotated the prism through successive angles of 90°. He said that the light was *polarised;* for, favouring the corpuscular theory, he concluded that the corpuscles possessed *poles* which gave rise to the observed effects. (The plane of reflection of the polarised light is called the *plane of polarisation.*) Extending his investigation, he found that the light which is reflected from the surface of any transparent medium, at definite angles (called the angles of polarisation) which depend upon the nature of the medium, exactly resembles one of the beams which have passed through a doubly refracting substance.

If the plane in which the light is reflected is parallel to the optic axis of the spar, the reflected light has the same properties as the ordinary ray in Iceland spar has; it manifests the properties of the extraordinary ray if its plane of reflection is perpendicular to the axis.

The supporters of the undulatory theory at first regarded the vibrations as taking place in the direction in which the waves travelled, but the phenomena of polarisation cannot be explained on this assumption. In particular, the conditions which are essential to the production of interference of polarised light (§ 221) necessitate the assumption that the vibrations take place perpendicularly to the direction of the ray.

Hooke, in 1672, had suggested that the vibrations occur in directions which are perpendicular to the ray; but the idea was never developed until its truth was inferred by Young and Fresnel, independently, not long after Malus had discovered that light was capable of undergoing polarisation by reflection.

A ray of light in which the vibrations of the ether all take place in one common direction, evidently possesses '*form*' or '*disposition*,' or '*sides*.' Think, for example, of a stretched cord placed between two smooth parallel planes which just touch it. Waves in which the vibrations are parallel to these planes can pass along the cord; perpendicular vibrations are incapable of existing. The waves possess '*sides*' which are in, and perpendicular to, the planes which confine the cord.

212.—*Laws of Polarisation by Reflection and Refraction.*—(1) *Brewster's Law.* Brewster made an elaborate series of investigations on the angles of polarisation of various substances, with the object of connecting the phenomenon with other optical properties of the substances. He found that *the index of refraction is equal to the tangent of the angle of polarisation.*

From this law we can at once deduce the relation $\cos i = \sin r$, where i and r are the angles of incidence and refraction. Hence, *the refracted ray is perpendicular to the reflected ray.*

Since the refractive index varies with the wave-lengths, rays of different colours are polarised at different angles.

Jamin has found that the polarisation is not quite complete except in some substances the refractive index of which is about 1·46.

(2) *Arago's Law.* The light which is refracted into a transparent medium is polarised to a greater or less extent. Arago found that *the quantity of polarised light in the refracted beam is equal to the quantity in the reflected beam, and the planes of polarisation of the two are at right angles to each other.*

Brewster's Law is applicable to reflection in the interior of a dense substance. Hence part of the unpolarised light in the refracted beam will undergo further polarisation when it is reflected at the second surface of the substance. If a sufficient number of parallel reflecting surfaces, such as those of a number of thin plates of glass placed one behind the other, be provided, the incident light may be divided into a reflected and a refracted beam, each of which is totally polarised in a plane at right angles to the plane of polarisation of the other. This arrangement constitutes a 'pile of plates.'

(3) *Malus's Law.* Light, which is incident at the polarising angle on a plane reflecting surface, is totally unaffected, as regards intensity, by a second reflection from a parallel plate of the same substance. But, if the plane of incidence upon the second plate be perpendicular to the plane of reflection from the first, the reflected beam will be totally extinguished. This subject was fully investigated by Malus, who found that *the intensity of the twice-reflected beam is proportional to the square of the cosine of the angle of inclination of the two planes of reflection.*

213. *Direction of Vibration in Polarised Light.*—The reflected ray, which (according to definition) is polarised in the plane of reflection, has symmetry with regard to that plane, since its intensity is totally unaltered by any number of reflections in that plane. It has also symmetry with regard to the normal to the plane of reflection, since it vanishes on reflection in any plane which passes through this normal.

We may therefore assume either that the direction of the vibrations in polarised light is perpendicular to the plane of reflection, or that it lies in the plane of reflection. Fresnel, in his theoretical investigations, made the former assumption; Maccullagh and Neumann adopted the latter.

The truth of the former is indicated by a number of considerations.

The vibrations of the ordinary ray in Iceland spar will be perpendicular to the optic axis provided that the vibrations of a ray polarised by reflection are perpendicular to the plane of polarisation;

and thus the uniform speed of that ray, in all directions, is readily accounted for. But, on the alternative assumption, the properties of the ordinary ray would be exceedingly difficult of explanation.

The ordinary and extraordinary rays produced by transmission through certain crystals, such as tourmaline, are coloured. When the two rays pursue nearly the same paths, identical colours are exhibited in each; and this occurs when the two rays traverse the substance nearly in the direction of the optic axis, so that their vibrations are nearly perpendicular to it. As the rays separate out from the axis, the colour of the ordinary ray remains constant, while that of the extraordinary changes greatly. Haidinger remarked that this favours the assumption that the vibrations of the ordinary ray are normal to the optic axis, and therefore take place along the normal to the plane of polarisation.

Carvallo has shown that the variations in the value of the coefficient x in Briot's formula of dispersion (§ 188), with a varying angle of incidence, indicate that Fresnel's assumption is correct.

If a horizontal beam of polarised light, the vibrations of which are in lines inclined at an angle α to the vertical, falls perpendicularly on a diffraction grating, the lines of which are vertical, the direction of vibration in the diffracted beam will make with the vertical an angle β which differs from α. Let a be the amplitude of the incident vibration. The resolved part of it parallel to the lines of the grating is $a \cos \alpha$; and the part at right angles to this is $a \sin \alpha$. If the diffracted beam makes an angle ϕ with the normal to the grating, the part of $a \sin \alpha$, which is perpendicular to the diffracted beam, is $a \sin \alpha \cos \phi$, and it is this part alone which is effective in the propagation of light. Hence the tangent of the angle which the new direction of vibration makes with the lines of the grating is $\tan \beta = a \sin \alpha \cos \phi / a \cos \alpha = \tan \alpha \cos \phi$. The angle β is therefore less than α. Consequently, if the plane of polarisation is perpendicular to the direction of vibration, the plane of polarisation of the diffracted beam will be more nearly perpendicular to the lines of the grating than that of the incident beam; and the reverse will happen if the direction of vibration lies in the plane of polarisation.

This result was deduced from theory by Stokes. He also tested it experimentally, and found that the result seemed to support Fresnel's assumption.

Another test, also due to Stokes, is based upon the nature of the polarisation of light which has undergone reflection from very small material particles.

Stokes remarks that no conclusion can be drawn so long as the particles are large compared with the wave-length of light, for then reflection occurs as it would from the surface of a large solid; but when the particles are small compared with the wave-length, it seems plain that the vibrations in the incident and the reflected rays cannot be at right angles to each other.

The small particles with which he experimented were obtained by highly diluting some tincture of turmeric with alcohol and adding water. A horizontal beam of sunlight fell upon the particles, and the light was found to be polarised in the plane of reflection. The smaller the particles were, the greater was the tendency to complete polarisation in the plane of reflection.

Since the 'sides' of the reflected ray are symmetrical with respect to the plane of polarisation, its vibrations must either be parallel to the incident ray or perpendicular to the plane of reflection, *i.e.*, of polarisation. We must therefore choose the latter alternative, since we cannot suppose that the directions of vibration in the incident and the reflected rays are at right angles to each other.

214. *Reflection and Refraction of Polarised Light.*—Young first determined the relations existing amongst the intensities of the incident, the reflected, and the refracted beams when light falls perpendicularly upon the bounding surface of two transparent media.

Fresnel, starting from certain assumptions, gave a complete investigation of these relations for all angles of incidence. The results have been experimentally verified.

He assumed, first, the conservation of *vis viva* (or energy); second, continuity of displacement of the particles of the ether at either side of the bounding surface; third, proportionality of the density of the ether in a given medium to the square of the refractive index of that medium.

The third assumption implies that the rigidity of the ether (regarded as possessing properties analogous to those of an elastic solid) is the same in any two media. For the refractive indices are inversely as the speeds of propagation of light in the two media, and therefore the densities are inversely as the squares of the speeds. But (§ 56) the squares of the speeds are in direct proportion to the ratios of the rigidity to the density of each medium, from which it follows that the rigidity of the ether in each must be the same.

At perpendicular incidence, the plane of polarisation is not changed by reflection. As the angle of incidence increases, the plane of polarisation becomes more and more nearly coincident with the plane of incidence. At the polarising angle, the two coincide entirely; and, at larger angles of incidence, the plane of polarisation passes to the other side of the plane of incidence—the angle between the two increasing until, at grazing incidence, the original and final planes of polarisation are equally inclined to the plane of incidence on opposite sides.

So long as the polarising angle is not exceeded, the difference of phase of the two components of the reflected vibration, respectively parallel and perpendicular to the plane of incidence, is zero; at the polarising angle the difference changes suddenly to π.

The plane of polarisation is not changed by refraction at perpendicular incidence. As the angle of incidence increases, it is more and more turned away from the plane of incidence. The rotation is doubled by refraction *through* a parallel plate.

The continuity of displacement contemplated in Fresnel's second assumption is continuity of tangential displacement. No account has been taken of the displacement perpendicular to the surface. If we replace Fresnel's assumption of uniform rigidity by the condition of no normal discontinuity, we find that the density of the ether must be the same in all media. Hence Fresnel's third assumption is inconsistent with normal continuity of displacement.

Making the assumption that the ether is of uniform density in all media, Macculagh and Neumann deduced Fresnel's laws for the reflection of polarised light provided that it be assumed that the vibrations are *in* the plane of polarisation. Though the result of this theory differs from that of Fresnel's in the case of a *single* refraction, the difference occurs in such a way that the result, after refraction *through* a parallel plate, is identical on both theories. And, further, the rotations of the planes of polarisation are of the same magnitude and sense on the two theories. Therefore none of the phenomena with which we are now dealing are capable of furnishing a test between the assumptions of uniform density and uniform rigidity.

215. *Plane, Circular, and Elliptic Polarisation.*—In the special examples considered in last section, the difference of phase of the two rectangular components of the resultant vibration was 0 or π. But the resultant of two rectangular simple harmonic motions is, in general (§ 34), elliptic motion. Not only is this true of two rectangular components; it is true of any number of simple harmonic components in lines inclined at any angles to each other.

Hence, *if we can assume that the vibrations of a particle of the ether are simply harmonic when polarised light (such as we have hitherto considered) is passing*, we must conclude that the most general vibration of such a particle, when subject to various simultaneous disturbances, is elliptical. This assumption is justified by the fact that no phenomena of light, which are not due to simple superposition of displacements, are observed.

The resultant elliptic path is described continuously so long as the amplitudes, phases, and periods of the components remain constant. In this case the light is said to be *elliptically polarised.* As a particular case, when all the components can be compounded into two rectangular components, equal in amplitude and period, but differing in phase by $\pi/2$, the ellipse becomes a circle, and the light is *circularly polarised.*

Ordinary polarisation—for example, that produced by reflection—occurs when the components can be reduced to two which differ in phase by any multiple of π. This is usually termed *plane polarisation,* in order to distinguish it from the above forms.

216. *Nature of Common Light.*—Common light exhibits no trace of polarisation of any description. But this is known to be true also of plane polarised light if its plane of polarisation be made to rotate very rapidly—so rapidly that, in little more than one-tenth of a second, the directions of vibration have been practically distributed uniformly in all possible orientations perpendicular to the ray. Hence we may conclude that ordinary light consists of elliptically polarised light, the magnitude, form, and position of the ellipse being in a constant state of rapid change.

But the phenomena of interference of light show that practically no change occurs in the course of some thousands of vibrations, for many thousands of interference bands can be counted when homogeneous light is used. On the other hand, since light travels at the rate of 186,000 miles per second, while the length of a wave is, on the average, about one forty-thousandth part of an inch, many millions of millions of vibrations must take place per second. But, again, as Stokes has pointed out, from the facts that every source of common light consists of a practically infinite number of points, and that the light emanating from each of these points is, in general, totally independent of that issuing from any other in respect of direction of vibration and also in respect of phase, we cannot expect anything else than an average effect in which there is no manifestation of 'sides.'

It follows that a beam of common light must necessarily be divided into two beams of equal intensity when it is transmitted through a doubly refracting substance.

Fresnel's theory shows that, in light reflected from the surface of a transparent substance, there is an excess of light polarised in the plane of incidence over that polarised in the perpendicular plane. It also shows that the reflected light is entirely polarised in the plane of incidence at the polarising angle.

The theory shows also that the refracted beam contains an excess of light polarised perpendicularly to the plane of incidence, that it is entirely polarised in the perpendicular plane at the polarising angle, and that its intensity is then equal to that of the reflected beam. At all angles of incidence there are equal amounts of polarised light in the two beams.

The known laws of the polarisation of common light by reflection and refraction are therefore consequences of the undulatory theory.

217. *Metallic Reflection.*—Malus observed that light is never completely polarised by reflection from the surface of metals, but that the polarisation attained a maximum at a certain angle of incidence. He also observed that polarised light appeared to be completely depolarised by reflection from a metallic surface when its plane of polarisation was inclined at an angle of 45° to the plane of incidence.

Brewster verified, and extended, these results. He showed that the reflected portion of a beam of common light might be completely polarised by a sufficient number of reflections under like conditions—a result previously inferred by Biot. He found also that, when the incident ray is polarised in, or perpendicular to, the plane of incidence, the reflected ray is still polarised in the same plane; that, when the original polarisation is in any plane other than these, partial depolarisation seems to take place; and that the depolarisation is greatest at the angle of maximum polarisation. Further, a second reflection, in the same plane, and at the same angle, repolarises the light; and the new plane of polarisation lies on the opposite side of the plane of incidence and makes a different angle with it.

The 'depolarisation' above spoken of does not mean restoration to the condition of common light. The originally polarised light may be decomposed into two parts, one polarised in the plane of incidence, the other polarised in the perpendicular plane. The amplitudes of these parts may suffer change by reflection, which (§ 214) produces a rotation of the plane of polarisation. The phases may also be altered, and this will give rise to elliptic polarisation. Jamin's experiments on this subject show that Brewster's 'depolarisation' is really elliptical polarisation, and also show the nature of the variations of amplitude and phase.

The laws of change of amplitude are the same as those given by Fresnel's theory. The difference of phase increases from perpendicular incidence to grazing incidence by the total amount π, the phase of the ray polarised in the plane of incidence being accelerated with reference to that of the other. The change of phase is extremely slow, except in the immediate neighbourhood of the angle of maximum polarisation, between near limits on either side of which practically all the change occurs.

The difference of phase ought, according to Fresnel's theory (§ 214), to increase suddenly by π at the polarising angle.

Extending his observations to transparent bodies, Jamin found that the difference between them and metals is only a difference of degree.

In all cases elliptic polarisation is produced, and the maximum ellipticity occurs at the angle of maximum polarisation, which coincides very closely with the angle deduced from Brewster's law.

Jamin found that some transparent substances differ from metals with respect to the sign of the difference of phase which is produced by reflection. Substances whose refractive index is less than 1·46 *retard* the phase of the component which is polarised in the plane of incidence; substances which have a refractive index exceeding 1·46 resemble metals in accelerating the phase of this component; substances in which the refractive index is equal to 1·46 obey Fresnel's laws.

He also found that, in metals, the angle of maximum polarisation decreases as the wave-length of the light increases; from which we see that metals, if they obey Brewster's law, must refract light of long wave-length more than light of short wave-length. Kundt's recent experiments on refraction through thin metallic prisms seem to confirm this conclusion.

When light polarised in the plane of incidence is reflected from a metallic surface, the intensity of the reflected beam is a minimum at the angle of maximum polarisation. Maccullagh pointed out that transparent substances, the refractive index of which exceeds $2+\sqrt{3}$, possess (according to his theory) a minimum reflecting power at a definite angle of incidence.

218. *Double Refraction by Biaxal Crystals.* — Brewster discovered that most doubly-refracting crystals possess *two* optic axes. In uniaxal crystals the axis is equally inclined to the three edges which meet at an obtuse-angled corner of the crystal. In biaxal crystals the lines which bisect the two angles contained by the axes, and the line at right angles to these two, have a definite relation to the crystalline form.

Fresnel has proved, theoretically and experimentally, that neither of the two rays in a biaxal crystal obeys the ordinary law of refraction. By means of certain assumptions, he investigated the problem of the propagation of waves of transverse vibration in a non-isotropic elastic medium. Crystalline substances are known to be non-isotropic, and presumably the property is impressed upon the ether which pervades them, and which is known to be hampered, as regards its free oscillation, by the presence of material particles.

The complete laws of double refraction may most readily be studied from the point of view of his theory, which we now proceed to consider.

219. *Fresnel's Theory of Double Refraction.*—Fresnel undertook his investigation when the discovery of double refraction in biaxal crystals made it apparent that Huyghens' construction for the wave-surface was not applicable in all cases.

In a non-isotropic substance, the resultant force which opposes the displacement of a particle does not in general act in the direction of the displacement. But Fresnel showed that there are three directions, at right angles to each other, in which the force, called into existence by the displacement, acts so as to move the particle directly back to its position of equilibrium. He showed that if, from any point in the interior of the substance, lines be drawn with lengths proportional to the square roots of the elastic forces which resist displacement in the directions in which the lines are taken, the extremities of these lines will lie on an ellipsoid (called the ellipsoid of elasticity). The three principal axes of this surface are in the directions in which the force tends to move the displaced particle directly back to its undisturbed position.

The speed of wave-propagation in an elastic medium is proportional (§ 56) to the square root of the elastic force (or distortional rigidity), and hence the radii of the ellipsoid are proportional to the speeds of propagation of waves when the vibrations are along the given radii.

Consider a plane-wave passing through the medium. Fresnel proved, from the fact that the intensity of a beam of light which is compounded of two beams polarised at right angles to each other is independent of the phase of either component, that the vibrations must lie in the wave-front. If, therefore, we regard a central section of the ellipsoid of elasticity by the wave-front, we see that there are only two directions—those of the two axes of the section—in which a displacement will give rise to a reverse force acting in a plane which is normal to the wave and which passes through the line of displacement; for Fresnel showed that the force acts in the normal to that central section of the ellipsoid which is conjugate to the direction of the displacement. In general, the force will have a component perpendicular to the wave-front; but, according to assumption, this produces no effect in the way of wave-propagation.

Corresponding to any given plane wave-front, there are therefore only two directions of vibration such that the elastic force developed by the displacement has an effective component entirely in the direction of the displacement. But this condition is *essential* to the propagation of a permanent wave. Hence a plane-wave, incident upon such a medium, is, in general, broken up into two waves, which are propagated in different directions with speeds which are proportional to the radii of the ellipsoid drawn in these directions.

[The following extract from Stokes' 'Lectures on Light' will aid in the formation of clear ideas on this point:

'Now we have not far to go to find a mechanical illustration of such a mode of action. Imagine an elastic rod terminated at one end, and extending indefinitely in the other direction. Let the rod be rectangular in section, the sides of the rectangle being unequal, so that the rod is stiffer to resist flexure in one of its principal planes than the other. Let this rod be joined on to a cylindrical rod forming a continuation of it which extends indefinitely. Conceive the compound rod as capable of propagating small transverse disturbances, in which the axis of the rod suffers flexure. Imagine a small disturbance, suppose periodic, to be travelling in the cylindrical rod towards the junction. It will travel on without change of type, even though the flexure of the axis be not in one plane. But to find what disturbance it excites in the rectangular rod, we must resolve the disturbance in the cylindrical rod into its components in the principal planes of the rectangular rod, and consider them separately. Each will give rise in the rectangular rod to a disturbance in its own plane, but the two will travel along the rod with different velocities. This illustrates the subdivision of a beam of common light falling on a block of Iceland spar into two beams polarised in rectangular planes, which are propagated in the spar with different velocities. Again, suppose the original disturbance in the cylindrical rod confined to one plane. If this be either of the principal planes of the rectangular rod, the more slowly or the more quickly travelling kind of disturbance, as the case may be, will alone be excited in the latter; and if the plane of the original disturbance be any other, the components into which we must resolve it in order to find the disturbance excited in the rectangular rod will in general be of unequal intensity, their squares varying with the azimuth of the plane of the original disturbance in accordance with Malus's law. This illustrates the subdivision of a beam of polarised light incident on Iceland spar into two of unequal intensity polarised in rectangular planes, and their alternate disappearance at every quarter of a turn. We see with what perfect simplicity the theory of transverse vibrations falls in with the elementary facts of polarisation discovered by Huyghens, standing in marked contrast in this respect with the conjecture by which Huyghens himself attempted to account for double refraction.']

In any one direction in the interior of the substance two plane waves may be propagated, generally with different velocities; and the vibrations in these waves are necessarily at right angles to each other. But there are two directions in which the speed of propagation of a wave is independent of the direction of vibration. These directions are parallel to the normals to the two sets of circular sections of the ellipsoid of elasticity; for, all the radii of a circular section being equal in length, the same force of restitution is called into play by a given displacement along any radius.

The resolution of an incident beam into two rectangularly polarised beams which obey Malus's law, and the existence of two directions in which a polarised beam is transmitted without modification, are therefore consequences of Fresnel's theory.

The planes of polarisation pass through the normal to the wave and through the major and minor axes of the section of the ellipsoid of elasticity by the wave. But the lines in which the circular sections cut the given plane section are equally inclined to the principal axes. Hence *the planes of polarisation bisect the dihedral angles which are contained by the planes which pass through the normal to the wave and the optic axes.*

The form of a wave which spreads out through the medium from any centre is found by the following construction, which is due to Fresnel: Along the normals to any central plane section of an ellipsoid, similar to the ellipsoid of elasticity, measure, from the centre, lengths which are proportional to the principal axes of the section. The wave-surface is the locus of the points so found, and consists of two sheets. The directions of the refracted rays are found from this surface by Huyghens' construction. The tangent plane drawn to one sheet gives the direction of the one ray; that drawn to the other sheet gives the direction of the second ray.

The wave-surface possesses symmetry with respect to the three principal planes of the ellipsoid from which it is derived. Its traces upon these planes consist respectively of a circle and an ellipse. Let OA, OE, OF (Fig. 125) be the three principal axes. AB is a circle of radius OA equal to the mean principal axis of the ellipsoid; CD and EF are circles the radii of which are respectively equal to the least and the greatest principal axes. BC is an ellipse, the principal axes of which are equal to the two least axes of the ellipsoid, and so on.

When the ellipsoid of elasticity is one of revolution, the wave-surface consists of two separate sheets—a sphere and an ellipsoid—which have one axis in common. Hence Fresnel's theory explains the double refraction of uniaxal crystals.

It is very desirable to note that Fresnel's results, although they are all seemingly in complete accordance with observed facts, are not rigorous deductions from his assumptions. Green and Neumann proved that a strict investigation, based upon these assumptions, will lead only approximately to Fresnel's Laws. Green then deduced these laws as rigorous consequences of an originally more general theory, the generality of which was subsequently limited by imposing a very probable condition; but this theory made it necessary to suppose that the direction of vibration is in the plane of polarisation. He then showed that a still more general theory would, by means of suitable assumptions, lead to the same laws, and to the conclusion that the vibrations are perpendicular to the plane of polarisation. Other theories also give like results—for example, Macculagh's theory, which is identical in its results with Green's first theory. On this point Stokes remarks that the principle of transverse vibrations is common to all these theories, while Fresnel's Laws are the simplest which can suit the phenomena; so that their mutual agreement can merely be regarded as a confirmation of that principle, while none of the special assumptions made as to the nature of the luminiferous medium in the interior of

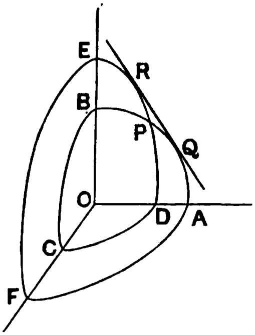


FIG. 124.

crystals can be regarded as being proved solely by the correctness of the results to which they lead.

220. *Conical Refraction.*—Sir W. R. Hamilton showed that the tangent plane QR (Fig. 124) touches the wave-surface at all points of a circle of contact, so that the point P is a 'conical point.' Four such points, one in each of the quadrants in the plane EOA, exist. They are, of course, the four points of intersection of the circle APB with the ellipse DPE.

The line OQ is perpendicular to the plane RQ. But the perpendicular on the plane which touches the wave-surface represents the speed of propagation of the plane wave to which it is a normal; and the direction of the rays in the crystal are those of the lines joining O to the point of contact. Hence a plane wave, incident upon the crystal in such a direction that RQ is its front after refraction, gives rise to a cone of rays which proceed in the directions of the lines joining O to the various points of the circle of contact. Hamilton's theoretical prediction of this phenomenon, which is known as *internal conical refraction*, was verified experimentally by Lloyd.

Both sheets of the surface are touched by the plane RQ, and hence there exists only one wave-velocity in the direction OQ, which is therefore one of the optic axes. The other optic axis is the image of OQ in the plane EOF.

An infinite number of tangent planes may be drawn to the surface at the point P, and therefore a ray which proceeds through the crystal in the direction of OP, gives rise on emergence to a thin conical sheet of rays. This *external conical refraction* was also predicted theoretically by Hamilton, and found, as the result of experiment, by Lloyd.

Since the radii of the wave-surface represent the velocities of the rays, we see that OP is a direction of single ray-velocity in the substance, for the two sheets of the wave-surface intersect at the point P. The other axis of single ray-velocity is the image of OP in EOF. The axes of single ray-velocity never deviate much from the axes of single wave-velocity, *i.e.*, from the optic axes.

221. *Interference of Polarised Light.*—The laws of interference of polarised light were investigated experimentally by Fresnel and Arago.

Two rays of light, which are polarised in perpendicular planes, do not give rise to phenomena of interference under circumstances in which rays of ordinary light would interfere.

The planes of polarisation of two such rays may, by suitable means, be made to coincide; but no interference will take place unless the two rays were originally parts of one polarised beam.

Such rays polarised in the same plane will always interfere under the conditions in which rays of ordinary light would interfere.

When rays which have been polarised by double refraction produce interference, the phenomena exhibited are such as to necessitate the assumption that the phase of one component has been accelerated by the amount π relatively to that of the other. The reason for this is not far to seek.

Let OP (Fig. 125) represent the vibration in a ray which falls upon a crystal of Iceland spar the principal section of which is AA′. OP will be broken up into two components O*n* and O*m*, respectively along and perpendicular to the axis. Let each of these be again resolved in the directions *aa′* and *bb′*. The components along *aa′* are in the same phase, but the components along *bb′* are necessarily in opposite phases. (Compare Fig. 126.)

222. *Colours of Crystalline Plates.*—Let us suppose that a beam of plane polarised white light is obtained by means of reflection, and let a second reflector be so placed as to extinguish the beam. If a thin crystalline plate be now interposed between the two reflectors, in the path of the beam, intensely coloured light will in general be reflected from the second surface.

The light disappears whenever the principal section of the crystalline plate coincides with, or is perpendicular to, the plane of reflection from the first surface. If the plate be turned round, in its own plane, from this position, some reflection from the second surface will be evident, and the reflected light will vary in intensity, but not in colour, as the plate is rotated. (The same appearances, *colour excepted*, would, of course, be manifested even if the plate were thick.) On the other hand, if the plane of reflection at the second surface be varied, the crystal being fixed, the colour

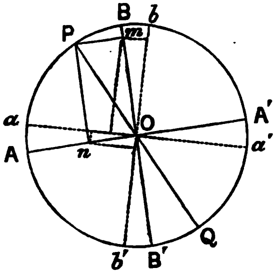


FIG. 125.

changes gradually into the tint which is complementary to the former.

When any two successive positions of the second plane of incidence differ by 90° from each other, the reflected tints are complementary.

The colours depend upon the thickness of the crystalline plate and vary with the thickness in the same way as the colours seen by reflection from a thin plate of air.

The explanation of the phenomenon is simple. The originally plane polarised light is divided in the crystal into two beams, which are oppositely polarised, and which traverse the plate with unequal speeds. The phase of the one beam is therefore retarded relatively to that of the other, and so interference may take place if the vibrations in the two portions are again resolved in a common direction. The tint is due to the cutting out of some rays, and the intensifying of others, by interference.

The light which emerges from the plate is elliptically polarised, since it is compounded of two rectangularly polarised beams which differ from each other in phase. In particular, when the difference of phase is any odd multiple of a quarter of a period, the light is circularly polarised; and, when the difference is a multiple of half of a period, the light is plane polarised. If the difference is an even multiple of a half-period, the plane of polarisation coincides with the original plane; it makes an equal angle with the principal section, on the opposite side, when the difference is an odd multiple of a half-period.

Let POP′ (Fig. 126) represent the direction of vibration in the incident beam of light which falls upon the plane surface POM of a doubly refracting plate, and let $\mu\mu'$ represent the principal section of the crystal so that the vibrations in the ordinary ray are in the

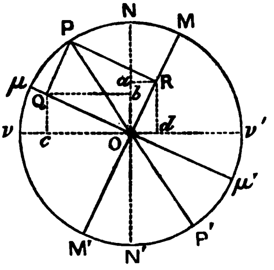


FIG. 126.

direction OM. If the amplitude of vibration along OP be unity, the amplitude of the vibrations in the ordinary ray is OR, and the amplitude of those in the extraordinary ray is OQ. Let $\nu\nu'$ represent the principal plane of a second doubly refracting crystal through which pass both the beams into which the original beam is divided.

The vibrations along OM and $O\mu$ give rise to two sets of vibrations along ON, the amplitudes of which are Oa and Ob respectively. These two vibrations are the components of the vibrations of the ordinary ray which emerges from the second crystal.

The vibrations along OM and $O\mu$ also give rise to two sets of vibrations along $O\nu$, the amplitudes of which are Od and Oc respectively. And these two vibrations are the components of the vibrations of the extraordinary ray which emerges from the second crystal.

If the vibrations Oa, Ob were in the same phase, the resultant amplitude would be $Oa+Ob$, and the intensity of the ordinary ray would therefore be $Oa^2+Ob^2+2\,Oa.Ob$. On the other hand, if the vibrations differ in phase, the resultant is (§ 34) represented by the third side of a triangle of which Oa and Ob form the remaining sides, the angle between them being the difference of phase. Hence the resultant intensity is $Oa^2+Ob^2+2kOa.Ob$, where k is the cosine of the difference of phase. From the diagram we see that we can write this as

$$1-Oc^2-Od^2+2kOa.Ob.$$

In the same way, we see that the intensity of the extraordinary ray is $Oc^2+Od^2-2kOc.Od$.

From the similar triangles QOb, ORa, we get $Oc.Od=Oa.Ob$, and so the intensity of the extraordinary ray is

$$Oc^2+Od^2-2kOa.Ob.$$

The difference of phase, and therefore k, depends upon the wave-length; and so, when light of various wave-lengths is used, we must take the sum of all such expressions.

From these expressions we can deduce all the observed effects.

The sum of the two intensities is unity because the colours of the ordinary and extraordinary rays are complementary.

The last term in each expression is the quantity upon which the colour depends. Hence all colour vanishes when Oa or Ob vanishes. This occurs when

$$POM=0 \text{ or } \frac{\pi}{2},$$

that is, when the principal section of the crystalline plate is parallel to, or perpendicular to, the original plane of polarisation.. It also occurs when

$$NOM=0 \text{ or } \frac{\pi}{2},$$

that is, when the principal sections of the thin plate and the second doubly refracting crystal are parallel or mutually perpendicular. The reason for the colours of the two rays being complementary is that the difference of the phases of the components of the extraordinary ray necessarily differs from that of the components of the ordinary ray by the amount π.

The colour is greatest when $Oa.Ob$ is a maximum. This takes place when $Oa=Ob$, which occurs when OP bisects μOM and coincides with, or is perpendicular to, ON. We then have

$$POM=\frac{\pi}{4}, PON=0 \text{ or } \frac{\pi}{2}:$$

that is, the principal section of the crystalline plate must make an angle of 45° with the original plane of polarisation, and the principal section of the second doubly refracting crystal must be parallel, or perpendicular, to that plane.

The colours change through regularly recurring cycles as the thickness of the plate increases by successive equal stages, for the difference of phase is proportional to the thickness.

223. *Special Cases.*—Hitherto we have assumed that the incident beam of light is parallel. We shall now suppose that a diverging beam of polarised light traverses a uniaxal crystalline plate the parallel plane faces of which are perpendicular to the axis.

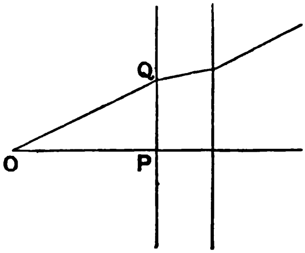


FIG. 127.

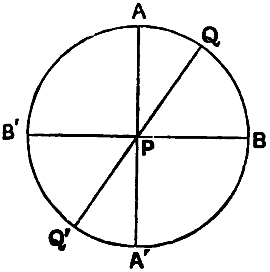


FIG. 128.

A ray OP (Fig. 127) which passes perpendicularly through the plate suffers no change. It will pass through, or be stopped by, a second plate which is cut parallel to its optic axis according as the principal section of the second plate is parallel to, or perpendicular to, the original plane of polarisation. Any other ray will undergo change according to its inclination to the optic axis, and according to the angle which the plane passing through it and the axis makes with the plane of polarisation.

Let AB′A′B (Fig. 128) be perpendicular to the axis of the cone of rays, and let AA′ and BB′ represent respectively the original plane of polarisation and the perpendicular plane. All rays which emerge from the crystal in these planes will be allowed to pass through, or will be stopped by, the second plate, according as its principal plane is parallel, or is perpendicular, to the original plane of polarisation. The field will therefore exhibit a light or a dark cross.

(*Haidinger's Brushes* are observed when polarised light is examined by the naked eye. The phenomenon consists of two yellowish-brown patches of light forming a brush the axis of which is parallel to the plane of polarisation; and two other bluish or purplish patches occur in the angles between the yellow patches. The appearance is due to a polarising structure which is most highly developed in dark eyes. It appears that the yellow spot of the eye is doubly refracting, and absorbs the extraordinary ray to a greater extent than the ordinary ray. Helmholtz found that the effect only appears with blue light. The brushes soon disappear unless the plane of polarisation be changed at intervals.)

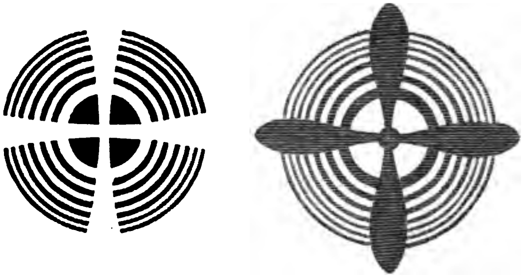


FIG. 129.

At any other point, such as Q, the vibration of the ray is resolved into its two components, polarised parallel and perpendicular to the plane through PQ. Thus the incident ray will be divided into two which traverse the crystal with different speeds and so give rise to interference. The retardation of the phase of the one component relatively to that of the other is constant so long as the distance PQ is constant, and becomes greater and greater as PQ increases in length. Hence the field exhibits a series of alternately light and dark circles surrounding the point P.

The circles are brilliantly coloured if white light be used; and the colours seen when the principal section of the second crystal occupies any definite position are exactly complementary to those which are seen when this plane is rotated through a right angle.

These effects are produced in comparatively thick crystals, since the difference between the speeds of the two rays, in directions not greatly different from the optic axis, is comparatively small.

The squares of the radii of successive circles are nearly proportional to the natural numbers. For it has been proved that the difference of the squares of the speeds of propagation of the two waves is proportional to the square of the sine of the angle which the ray within the crystal makes with the optic axis, and also that it is proportional to the thickness of the plate and to the interval of retardation conjointly. Hence the retardation varies as the square of the sine of the angle between the ray and the axis. But this angle is very nearly equal to the angle QOP (Fig. 127), or to the distance QP.

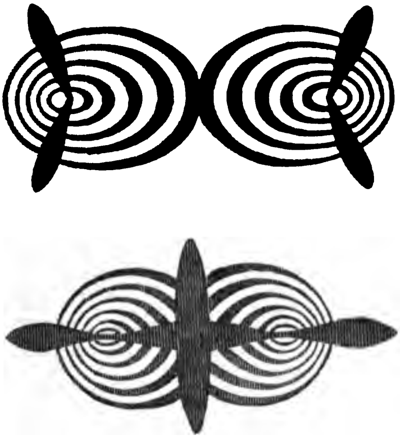


Fig. 130.

Consider now a parallel plate cut from a biaxal crystal in a direction perpendicular to the line bisecting the optic axes. The interval of retardation is, in this case, proportional to the product of the sines of the angles which the wave normal makes with the axes. And these sines are approximately proportional to the distances of the point of emergence of the ray from the points in which lines drawn parallel to the axes, from the point of incidence on the first face of the crystal, meet the second face. Hence the locus of P is an oval of Cassini.

224. *Artificial Production of the Doubly Refracting Structure.*—Fresnel showed that glass and other singly refracting substances become doubly refracting when subjected to stress; and Brewster showed that unequal heating of the substance is sufficient to produce the requisite strain. Compression gives rise to a structure which resembles that of Iceland spar in producing an extraordinary ray the refractive index of which is less than that of the ordinary ray. Expansion produces an opposite effect.

The optic axis, in these cases, is fixed in *position* as well as in direction, unless the strain be homogeneous.

If an elliptic cylinder of glass be suddenly heated uniformly over its surface, a biaxal refracting structure is developed. This structure is also usually seen in unannealed glass.

Analogous changes may be produced in substances which naturally exhibit double refraction.

Clerk-Maxwell showed that a viscous liquid becomes doubly refractive while subjected to shearing stress.

225. *Rotatory Polarisation.*—Quartz is a doubly refracting substance in which the refractive index of the extraordinary ray is greater than that of the ordinary ray. The wave-surface consists of a sphere and a spheroid, but the spheroid lies entirely within the sphere. It follows from this that the speeds of propagation of the two rays along the optic axes are not the same. But, further, the vibrations in the two rays are not rectilinear. They are elliptical in general, and the ellipses are described in opposite directions in the two rays. When the rays are propagated along the axis, the vibrations become circular.

Now, the resultant of two uniform motions, of equal period, in opposite directions in the same circle, is rectilinear motion. For, if A, A′ (Fig. 181) represent simultaneous positions of the oppositely moving points, the resolved parts of their motion perpendicular to the line PQ, which is drawn from the centre of the circle through the middle point of the arc joining A, A′, destroy each other. Hence the resultant is simple harmonic motion in the line PQ.

But, if A′ be retarded relatively to A, the line PQ will take a new position bisecting the arc between A and the new position of A′. And, if A′ is continuously retarded, PQ will revolve continuously round in the direction AA′. On the other hand, if A be retarded relative to A′, PQ will revolve in the direction from A′ to A.

The two circularly polarised rays in quartz therefore produce plane polarised light on emergence; but, because of the retardation of one ray relatively to the other, the plane of polarisation has been rotated through an amount which is proportional to the thickness of the quartz.

The rotation is right-handed in some specimens of quartz, left-handed in others. Amethyst consists of alternate layers of right-handed and left-handed quartz.

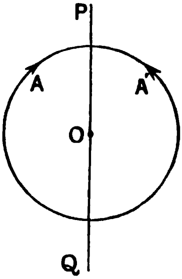


FIG. 131.

The amount of rotation is dependent upon the wave-length. Biot, and subsequently Broch, proved that it is approximately inversely proportional to the square of the wave-length. The first three terms of the formula

$$\rho=\frac{a}{\lambda^2}+\frac{b}{\lambda^4}+\frac{c}{\lambda^6}+\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

furnish a much better approximation, ρ being the rotation, and a, b, c being constants.

The spectrum produced by plane polarised sunlight which has passed through a plate of quartz is indistinguishable from that which is produced by ordinary sunlight. But a profound modification takes place if, previous to its passage through the refracting prism, the light be passed through an apparatus arranged to polarise light in a plane at right angles to the original plane of polarisation. Were the quartz plate taken away, no light could pass under these circumstances. The effect of the quartz is to restore the light, with the exception of such rays as have had their plane of polarisation rotated through a multiple of two right angles. Consequently, the spectrum is crossed by dark bands. By rotating the second polarising apparatus any one of the dark bands may be caused to occupy any desired position in the spectrum, and so the total rotation for any particular kind of light (and, consequently, the wave-length of that light) may be measured with extreme accuracy.

The dark bands are not sharply marked, for rays near those which are totally extinguished are necessarily partially extinguished. And, since portions of the light are cut out, the finally emergent light is coloured. The coloration disappears when the length of the quartz is so great that the portions which are cut out are very narrow and closely packed together.

Many liquids, solutions, and even vapours, possess this rotatory power—the rotation being in some cases left-handed, in others right-handed. As a rule, the rotation produced by a given thickness of a liquid is much smaller than that produced by the same thickness of quartz.

Neither dilution by an inactive substance nor vaporisation alters the rotatory power of a liquid, except in degree; on the other hand, Herschel showed that quartz is inactive when in solution; and Brewster showed that it is inactive when fused. Hence it is inferred that, while the rotation in quartz depends upon the crystalline structure, the rotation in liquids and vapours is a molecular phenomenon.

All rotatory polarimeters, or saccharimeters, depend essentially upon the above principles. In some of them, equal intensity of two beams of light is used as a test; in others, the test is furnished by the equality of tint of two beams (§ 226). A spectroscopic test, such as that which has been described above, is by far the most delicate.

If the light which has passed through quartz be reflected back, so as to retraverse it in the opposite direction, the rotation will be undone. Rotation of the plane of polarisation may be produced in a magnetic field; but, in this case, reversal of the path of the light will not (Chap. XXX.) be accompanied by an undoing of the rotation.

226. *Polarising Prisms.*—Many methods are used for the production of polarised light.

Plane polarisation by reflection has been already described.

Any doubly refracting crystal, of course, furnishes us with the means of producing plane polarised light, provided that we can separate the two beams.

Some substances, such as tourmaline, strongly absorb one of the two rays into which the incident beam is divided, and so furnish, when of sufficient thickness, a ready means of obtaining plane polarised light.

A block of Iceland spar separates the rays in proportion to its length, but it is extremely difficult to obtain large blocks which are free from internal flaws. In *Nicol's prism*, therefore, one of the two rays is got rid of by total reflection. A long block of the spar, whose parallel plane ends are cut at a suitable inclination, is divided into two parts by a plane which is perpendicular to its principal section. The two portions are then cemented together, in their original position, by means of Canada balsam, the refractive index of which is intermediate in magnitude between the indices of the spar for the extraordinary and the ordinary rays. Consequently, when the inclination of the dividing plane to the path of the ray is sufficiently great, the ordinary ray suffers total reflection at the surface of the film of balsam, and the extraordinary ray alone is transmitted.

Foucault's prism is essentially similar, but the film of balsam is replaced by a film of air. A shorter block of the spar suffices in this method, but considerable loss of light takes place by reflection at the surface of the film.

Considerable separation of the rays may be obtained by the use of a prism of Iceland spar, or of quartz, which is achromatised by means of a prism of glass. In *Rochon's prism* the edge, and one of the faces, of the prism are perpendicular to the optic axis. The rays therefore pass through the prism without separation until they emerge at the opposite face. The prism is achromatised by means of a second prism, of the same substance, the refracting edge of which is parallel to the axis. The ordinary ray proceeds through the second prism without alteration of direction, and is therefore uncoloured; but the extraordinary ray is considerably refracted, and is coloured at its edges, since the refraction depends upon the wavelength.

Greater angular separation of the rays may be obtained by *Wollaston's prism*. In this arrangement the refracting edge of the first prism is perpendicular to the optic axis, but the face upon which the light falls perpendicularly is parallel to the axis. In all other respects the arrangement is the same as in Rochon's prism. Both rays are coloured at their edges, for both are deviated from their original direction.

Such prisms may be used as analysers to determine the angular position of the plane of polarisation. Thus, when a Nicol's prism extinguishes polarised light, the plane of polarisation is parallel to the longitudinal principal plane of the prism. But it is not easy to judge by the eye the exact position of the prism at which extinction takes place, since the intensity of the illumination varies very slowly in the neighbourhood of a maximum or minimum value. It is much more easy to tell when two fields are equally illuminated, provided that the light be not too intense. This condition may be attained by the use of a compound prism the principal planes of the parts of which are slightly inclined to each other.

In Jellet's analyser, the ends of a block of Iceland spar are cut off by planes at right angles to the length. When light is incident perpendicularly on the end of such a prism, the ordinary ray passes straight through, while the extraordinary ray passes to one side. If the end edges are all equal, the longitudinal principal plane passes through the shorter end diagonal, and so the direction of vibration in the ordinary ray is parallel to the longer diagonal. If the prism were split longitudinally by a plane passing through the longer end diagonal, and if one of the halves were then reversed as regards its length and refitted to the other half, the direction of vibration in the ordinary ray would be unaltered. But if it were split by a plane slightly inclined to the longer end diagonal, and if one half were reversed, the directions of vibration in the ordinary rays in the two halves would be inclined to each other at an angle equal to twice the given inclination. When the plane of polarisation of the incident light bisects the exterior angle between these directions, the two halves are equally bright. The extraordinary rays in the two halves are oppositely diverted, and may be stopped out by means of a diaphragm.

In Laurent's analyser, the slight change of direction of vibration is produced by means of a parallel plate of quartz having one edge parallel to the optic axis. The thickness of the quartz is such that, with light of a given wave-length, that component of the incident vibration which is parallel to the optic axis is retarded, relatively to the other component, by half a wave-length. Hence, on emergence, the direction of vibration is inclined to the optic axis at the same angle as the incident vibration, but lies on the opposite side. The loss of light caused in the one half of the field by the quartz plate is compensated, in the other half, by means of a glass plate. Since, for a given thickness of quartz, the retardation varies with the wave-length, this prism can only be used with homogeneous light.

Poynting suggests the use of a parallel quartz plate cut perpendicular to the axis. One half is made slightly thinner than the other. Hence the plane of polarisation is rotated to a slightly less extent in the thin half than in the thick half (§ 225). In another modification of this method, also suggested by Poynting, a liquid which rotates the plane of polarisation is employed. It is placed in a glass vessel with parallel sides, and the effective thickness of one half is reduced by means of a plate of glass which is placed against one of the sides. As in the case of Laurent's analyser, homogeneous light must be used.

As we have seen, when non-homogeneous plane-polarised light is employed with a plate of quartz cut perpendicular to the axis, the light which emerges from the analyser is in general coloured. Some methods of determining the angular position of the plane of polarisation are based on this fact. If the thickness of the quartz plate be not greater than half a centimetre, the colour can be made nearly neutral. The light is distinctly coloured when the analyser is slightly rotated in one direction; it becomes as distinctly coloured with the complementary tint when the analyser is equally rotated in the opposite direction. Yellowish-green light is stopped by the apparatus when the neutral tint appears.

By the use of the *biquartz*, a still more accurate determination may be made. Light which has passed through a semicircular plate of right-handed quartz fills one half of the field of vision; light which has passed through an equally thick semicircular plate of left-handed quartz fills the other half. The colours of the two halves are in general different; but, if the thickness of the quartz plates be slightly greater than three millimetres, in one position of the analyser, the two halves take the neutral tint. The contrasted complementary colours which appear on a slight rotation of the analyser from this position furnish a fairly delicate test.

Methods of producing elliptically, or circularly, polarised light have been already indicated. They all depend upon the introduction of a difference of phase between the two rectangularly polarised components of a beam of light. This difference may be produced by transmission through a doubly refracting plate. When the plate is of such thickness as to produce a difference of phase of a quarter-period, the light is circularly polarised provided that the two beams are of equal intensity. Such a plate is termed a quarter-wave plate. The difference of phase may also be produced by reflection or refraction—as in *Fresnel's rhomb*. This consists of a parallelepiped of St. Gobain glass the faces of which are so inclined that a ray enters and emerges perpendicularly at opposite faces after two internal reflections at an angle of 54° 37′. Each reflection produces a difference of phase of 45°, and therefore the light emerges circularly polarised if it were originally polarised in a plane inclined at 45° to the plane of internal reflection.

Conversely, Fresnel's rhomb shows the existence of circularly polarised light by changing it into plane polarised light. This will also occur in the case of elliptically polarised light when either axis of the ellipse is inclined at 45° to the plane of internal reflection. But the two cases may be distinguished by means of a Nicol's prism; for rotation of the prism produces no alteration of intensity if the beam which is passing through it is circularly polarised, but it does produce variation of intensity if the beam is elliptically polarised. Though similar variation of intensity takes place if the incident beam is partially plane polarised, a quarter-wave plate, or a Fresnel's rhomb, enables us clearly to distinguish between these two cases.

CHAPTER XVIII.

THE NATURE OF HEAT.

227. *Radiant Heat. Its Identity with Light.*—We are accustomed to speak of the heat which we receive from the sun just as we speak of the light which we receive from it. And so the term 'radiant heat,' as applied to the heat which comes to us from distant bodies, apparently without the intervention of ordinary matter, has come into scientific use.

The non-intervention of ordinary matter in the process of radiation by which heat, like light, passes from one body to another at a distance from it, can readily be proved. There is no sufficient amount of ordinary matter in interstellar or interplanetary space to account for the transference; and a hot body cools in *vacuo* almost as readily as when surrounded by air—indeed, in certain cases, a hot body will cool less rapidly when surrounded by a material medium than it will otherwise.

Luminous bodies (those which exhibit fluorescence or phosphorescence being alone excepted) radiate heat as well as light, and the hotter they are the more luminous do they appear. We might, therefore, naturally conclude that the difference between light and radiant heat is not a difference of kind, but only a difference of degree. This inference is fully borne out by many facts.

One point in which they completely resemble each other is rectilinear propagation in free space or in a homogeneous medium. The heat-shadow which any obstacle casts is identical with the shadow which it produces with regard to light proceeding from the same source. And, since the path of light is rectilinear, this proves that radiant heat also moves in straight lines.

Heat, like light, is not propagated instantaneously; and an even more fundamental point of resemblance than the above appears in the fact that their speeds in *vacuo* are identical. This is made evident by the simultaneous disappearance and reappearance of the light and heat when a total eclipse of the sun takes place.

The laws of reflection of the two are identical, for the focus of a mirror for heat-rays is the same as its focus for light-rays. A thermo-electric pile, placed at the focus of a reflecting telescope, can make evident the heat radiated from a star.

Their laws of refraction are also the same; although, at first sight, a difference appears because the focus of a lens for heat is farther off from the lens than its focus for light. But this only strengthens the analogy, for, unless the lens be achromatised, the focus for red rays is farther off than the focus for blue rays.

Both are governed by the same laws of interference and of polarisation. The phenomena of interference *prove* the existence of periodicity of motion, and show that the vibrations are transverse, as in the case of light. And the usual methods based upon interference, diffraction, and refraction, enable us to measure the wave-length, which is found to be greater than that of luminous rays.

We therefore conclude that light and radiant heat are one and the same thing; that the latter differs from the former only as red light differs from blue light; and that it is not evident to the sense of sight because the eye is so constituted that it cannot respond to the slower vibrations. We already know that some eyes can perceive rays at the red end of the spectrum (and also at the blue end) which are totally invisible to other eyes.

And colours—the word being used by analogy—appear in heat just as they do in light. Rock-salt is very transparent to heat rays, just as glass is transparent to light; but, on the other hand, ordinary glass is very opaque to heat rays, *i.e.*, it absorbs them to a great extent. It acts to heat just as coloured glass acts to light; and many other substances act similarly. The law of absorption, with varying thickness of the medium, is the same as that which holds in the case of light (§ 179).

228. *Heat in Material Bodies. Hypothesis of Molecular Vortices.*—Since hot bodies give out radiation, and since the propagation of radiation involves motion of the particles of an inert medium, we might infer that the particles of a hot body must be in rapid motion, and that the communication of heat from one body to another depends upon the intercommunication of motion.

It is scarcely a century since Rumford and Davy arrived at this result upon experimental grounds.

Previous to their investigations heat was supposed to be a form of matter which was occluded in the interior of substances, but which was looked on as imponderable, since it did not add to the weight of a body. This imponderable substance was termed *Caloric*.

According to the caloric hypothesis, a body was hotter or colder in virtue of its having absorbed a greater or a smaller quantity of caloric; and when, by any means, the capacity of a body for caloric was diminished, heat (or rather caloric) was necessarily given out.

In the year 1798 Rumford was engaged in the boring of cannon, and had his attention forcibly drawn to the fact that there was a rise of temperature in the process of the reduction of the solid metal to the state of filings. But, according to the caloric hypothesis, the rise of temperature implies a diminution of the capacity of the substance for caloric; and, conversely, an increase of the capacity of a body for caloric would be accompanied by a fall of temperature unless additional caloric were supplied. Hence, in Rumford's experiments, the rise of temperature of the filings signifies, on this hypothesis, a diminution of the capacity for caloric.

Rumford, therefore, sought to determine by experiment whether or not the filings had less capacity for heat than the solid metal had. He heated equal weights of the solid metal and of the filings to the same high temperature, and dropped them into equal quantities of water at the same low temperature. He found that the same changes of temperature were produced in both cases, and concluded that the capacity of the substance for caloric had not been reduced when the body was broken up into smaller portions.

The difficulty might be got over by asserting that, although the capacity was not altered, the total amounts of heat contained in the equal masses in the different conditions were not equal—in other words, by asserting that the *total* capacity was altered. But, if the equal masses were dissolved, under identical conditions, in equal amounts of some acid, so as to reduce them to the same final state, this point might be tested. Still, there might be a distinct difference between the physical states of the two specimens of the substance which might cause this experiment to indicate a wrong result. For the filings might have been so strained as to contain a considerable amount of latent heat which would appear on their complete recovery from the state of strain. Rumford's conclusion was nevertheless correct, for it is entirely supported by experiments based upon accurate principles; and his observations therefore prove that the caloric hypothesis is incorrect.

But Rumford did not stop at this stage. He observed that the quantity of heat which was developed in the process of boring was independent of the amount of metal which was abraded—that a blunt borer and a sharp borer, though producing very different amounts of filings, caused the same development of heat if the same amount of work were spent in driving them. And, further, there seemed to be practically no limit to the amount of heat which might be produced. He therefore reasoned that heat could not be a material substance, and stated that he could scarcely conceive of anything but *motion* which could be excited and communicated in the manner observed.

Almost at the same time (in 1799) Davy was experimenting in precisely the same direction. He showed that two pieces of ice might be melted simply by rubbing them together. Now, heat is required to produce this change of state, and so, on the caloric hypothesis, the capacity of water for heat must be less than the capacity of ice for heat. But it is well known that the exact reverse is true.

The necessary heat might have been furnished by surrounding bodies, and therefore Davy enclosed the two pieces of ice by other portions of ice, and placed them in the exhausted receiver of an air-pump. Under these conditions heat could only reach them by first melting the surrounding ice.

Davy was entitled to conclude, from the result of his experiments, that heat was not a form of matter, but at the time he merely said, 'Friction, consequently, does not diminish the capacities of bodies for heat.'

In 1812, when again discussing the point, he spoke of heat as 'a peculiar motion, probably a vibration, of the corpuscles of bodies tending to separate them,' and said that 'The immediate cause of the phenomenon of heat, then, is motion, and the laws of its communication are precisely the same as the laws of communication of motion.'

The second statement is rigidly correct; the first—that heat is motion—is only true if properly interpreted. Heat, since it is not matter, must be *energy;* and so the true meaning of Davy's statement is that heat consists in the energy of motion of the particles of a material body. But the word 'energy' was not introduced into science at that time.

Davy illustrated this heat-motion, which he termed 'repulsive motion,' by the analogy of the orbital motion of the planets. If the speed of motion of any planet were increased, the orbit would become larger just as if a repulsive force had acted.

In Davy's statement we have therefore the complete foundation of the whole kinetic theory of gases (Chap. XI.), and, more generally, of the modern dynamical theory of heat.

Since heat is a form of energy we might infer the possibility of its existence in a potential form, and the use of the common term *latent* heat bears out the inference.

In working out a dynamical theory of heat Rankine advanced a hypothesis of 'molecular vortices.' He supposed that the motions which constitute heat in bodies are vortical motions in atmospheres around nuclei, and that radiation consists in the propagation of vibratory motions of the nuclei under their mutual forces. The energy of the vortices is the amount of heat which bodies possess; and the absolute temperature of any body is the quotient of this energy by a definite constant for each substance. The elastic pressure, according to dynamical laws, must be directly proportional to the vortical energy, and must be inversely proportional to the volume which the vortices occupy, except in so far as the mutual nuclear forces which exist in all non-perfect gases modify this result. Latent heat is the equivalent of work done in varying the dimensions of the vortical orbits when the volumes and shapes of the spaces which they occupy are altered. Specific heat is the equivalent of work spent in varying the vortical energy.

CHAPTER XIX.

RADIATION AND ABSORPTION OF HEAT.

229. *Prevost's Theory of Exchanges.* — The fact that the laws of communication of heat are precisely those of the communication of motion leads to the conclusion that motion of the particles of a hot body is not confined to the surface alone; and this conclusion is confirmed by the greater radiation which takes place from a thick plate, than from a thin plate, of a transparent substance, when both plates are at the same temperature. It indicates also that the radiation from a hot body is dependent upon the state of that body alone, and is not influenced by the presence of any other body, except in so far as it may cause an alteration in the thermal state of the former.

This is the essence of the Theory of Exchanges which was advanced by Prevost of Geneva, under the title of a theory of 'movable equilibrium of temperature.'

According to Prevost, two bodies, which are of different temperatures, and are placed in an enclosure which is impervious to heat, will both radiate heat. The hotter body will radiate at a greater rate than the other, until, by absorption, the temperatures of the two are equalized. After this, each will still radiate, but at precisely the same rate; so that the heat which one loses by radiation is balanced by that which it gains by absorption. This is the condition which Prevost termed a condition of movable equilibrium (or, as we would now call it, of kinetic equilibrium) of temperature.

230. *Stewart and Kirchhoff's Extension of Prevost's Theory.*— As in Chap. XV., we define the *absorptive power* of a body, under given conditions, for any definite radiation, as the fraction of the whole incident radiation of that kind which it absorbs; and we also define the *emissivity* of a body, at a given temperature, for any

given radiation, as the ratio of the quantity of that radiation which it emits to the quantity of it which is emitted by a black body under the same conditions.

And Stewart's proof, as given in Chap. XV., leads to the result that the *emissivity and the absorptive power of a body, at a given temperature, for any radiation, are equal.*

It is needless to enter into any discussion of special cases further than those which have already been given (§ 177). Suffice it to say that, whatever be the quality and the quantity of the radiation emitted by any body under given conditions, absorption must exactly balance emission, in respect both of quality and quantity, if the given conditions are to be maintained. Previously to Stewart's investigations, it was known from the experiments of Leslie, De la Provostaye, and Desains, that the radiating and the absorbing powers of any one body were to a considerable extent proportional to each other; that is to say, it was known that a good radiator was a good absorber, and that a bad radiator was deficient in absorbing power.

231. *Laws of Radiation of Heat at Constant Temperature.*—Early in the history of the subject, it was known that the radiation from a body at a given temperature depended upon the nature of the surface of that body. (This furnishes an additional analogy between light and radiant heat.)

In illustration of this fact Leslie constructed a hollow metal tube, one side of which was polished, while another was rough. A third side was covered with lampblack, and the fourth was coated with white enamel. Although the surfaces of the two latter sides were so very different, Leslie found that both radiated about equally well when the cube was filled with hot water. The radiation from the bright metallic surface was much smaller than that from any of the others, and the radiation from the rough metallic surface was considerably less than that from the enamelled and the blackened surfaces.

It has already been proved (§ 179) that the amount of any given radiation, which is emitted from a sufficient thickness of a substance of given radiating power, is equal to that which is emitted from a black body at the same temperature. It was shown that the amount which is transmitted through a plate of the substance, the thickness of which is n units, is $R(1-\rho)^n$, where the quantity R represents the total amount of the incident radiation and ρ is the absorption co-efficient. Hence the amount which is stopped by the plate is

$$R(1-(1-\rho)^n);$$

and therefore, by definition, the absorptive power—and, consequently, the emissivity—is

$$1-(1-\rho)^n.$$

As the temperature of a body rises, radiations of shorter and shorter wave-length are emitted, and the energy in each previously existing kind is increased.

232. *Heat Spectra. Diathermancy.*—If the radiation from a luminous body be passed through a slit and a prism in the usual manner, a spectrum is obtained which enables us at once to discover the nature of the radiation. And we may measure the amount of energy in any given portion of the spectrum by allowing that radiation to fall upon a medium which entirely absorbs it, and is consequently raised in temperature to a measurable extent. But, in any such experiment, it is necessary first to make certain that absorption does not take place, to any appreciable extent, in the substance of the prism.

We might also indirectly analyse the radiation emitted by a given body by means of determinations of the absorption which the body exercises upon radiations of different wave-lengths; but such measurements are of little value unless the substance used is of definite chemical composition and physical structure. Some gases (*e.g.*, olefiant gas) exert powerful absorption on the heat rays; others exhibit very little. The absorption produced by water vapour seems to be largely due to the dust nuclei (§ 251).

These remarks apply to the invisible portions of the spectrum also, whether these consist of rays of higher refrangibility, or of rays of lower refrangibility, than those which form the visible parts.

It is found that the invisible portions of the spectrum possess characteristics which are precisely analogous to those which appear in the visible parts.

De la Roche was the first to investigate the selective absorption of heat. His results were confirmed by the elaborate researches of Melloni. Some of the results obtained by the latter are given in the following table. The numbers give the percentages of the incident radiation which is transmitted through layers of various substances 2·6 mm. in thickness. The numbers in the first column were obtained when the source of heat was a Locatelli lamp without a glass chimney; those in the second refer to incandescent platinum heated by a spirit lamp as the source; and those in the third and fourth columns refer to blackened copper heated to 390° C. and 100° C. respectively. The result depends essentially upon the nature of the source except in the case of rock salt:

Rock salt (clear)	92	92	92	92
Rock salt (dull)	65	65	65	65
Fluate of lime (clear) ...	78	69	42	33
Fluate of lime (greenish) ...	46	38	24	20
Fluate of lime (green) ...	8	6	4	3
Iceland spar	39	28	6	0
Rock crystal	38	28	6	0
Mirror glass	39	24	6	0
Ice (very pure)	6	0	0	0

Some of his results for liquids are given below. The source was an Argand lamp with a glass chimney, and the thickness of the substances was 9·21 mm.:

Chloride of sulphur	63
Essence of turpentine...	31
Oil of olives	30
Naphtha (natural)	28
Naphtha (rectified)	26
Sulphuric ether...	21
Pure sulphuric acid	17
Nitric acid (pure, colourless)	15
Alcohol (absolute, colourless)...	15
Acetic acid (rectified)	12
Salt water	12
Distilled water	11

The same thickness of mirror glass transmitted, with this source, 53 per cent. of the incident radiation.

Melloni also proved that the absorption exercised by a given substance depends upon the absorption to which the radiation has been already subjected. As in the first table above, the numbers in the first column following refer to radiation from a Locatelli lamp and to a plate of the substance 2·6 mm. thick. The numbers in the second, third, and fourth columns respectively refer to the absorption when the radiation has already passed through a plate, 2·6 mm. thick, of alum, sulphate of lime, and chromate of potash respectively:

Rock salt	92	92	92	92
Fluor spar	78	90	91	88
Iceland spar	39	91	89	56
Rock crystal	38	91	85	52
Glass (white)...	40	90	83	50
Glass (opaque black) ...	16	0·5	18	11

Tyndall was the first to make an elaborate investigation of the diathermancy of gases and vapours. Some of his results for various gases, at a pressure of one atmosphere, the absorptive power of air being taken as unity, are here given:

Air	1	Carbonic acid ...	90
Oxygen	1	Nitrous oxide ...	355
Nitrogen	1	Marsh gas ...	403
Hydrogen	1	Olefiant gas ...	970
Carbonic oxide ...	90	Ammonia	1195

In the case of vapours, he finds, for example:

Bisulphide of carbon	15	47	62
Chloroform	85	182	236
Sulphuric ether	300	710	870
Alcohol	325	622	—
Acetic ether	590	980	1195
Boracic acid	620	—	—

The numbers in the first, second, and third columns give respectively the absorptions at pressures of 2·54, 12·7, and 25·4 mm. of mercury.

Tyndall also found that the absorptive powers of vapours diminish greatly as the temperature of the source increases—the source being a spiral of platinum at temperatures between that of bare visibility and that of fusion. He showed, further, that the order of vapours as regards absorption is the same as that of the corresponding liquids.

Great diversity of opinion exists regarding the absorptive power of water vapour. Tyndall found it to be excessively great in comparison with that of dry air. Magnus found that the two were large, and about equal. Lecher and Pernter found that they were both excessively small.

We cannot, by means of the thermopile (Chap. XXV.) the bolometer (Chap. XXVI.), or the radio-micrometer (Chap. XXV.), which are the most suitable instruments for the present purpose, determine the energy of the radiation of one definite wave-length. All that can be done is to measure the energy of the total radiation which is contained between rays of known wave-length, for the face of the thermopile and the metal strips of the bolometer and of the radio-micrometer necessarily possess considerable breadth.

In the case of a refraction spectrum, besides the difficulty regarding absorption by the substance of the prism, there is the difficulty of the crowding together of the rays towards the less refrangible end of the spectrum according to an unknown law.

If a diffraction spectrum be used, absorption may take place in the substance of the grating if it be made of glass; and, if it be a metallic grating, great uncertainty exists with regard to the nature of its action upon the invisible rays. And, although the dispersion be (§ 207) practically proportional to the wave-length, so that equal breadths of the spectrum correspond to equal differences of wave-length, we must remember that the measuring instruments ought to determine the energy contained in portions of the spectrum which are bounded by rays the wave-lengths of which are in a constant ratio. Further, since we are dealing with radiations of *all* wave-lengths, we see that, at any one part of a diffraction spectrum, an infinite number of different radiations are superposed, because of the existence of an infinite number of spectra of different orders.

The difficulties of the problem have been largely overcome by Langley, to whom is chiefly due our recent great increase of knowledge regarding radiations of large wave-length. He has detected, by means of the bolometer, traces of heat in portions of the solar spectrum corresponding to wave-lengths about twenty-four times greater than those of the least refrangible part of the visible spectrum.

He has shown that the wave-length at which the maximum of energy in the spectrum exists diminishes as the temperature is raised. This result has been deduced from theoretical considerations by Michelson; and the numerical deductions from the theory accord very well with Langley's observations. The energy at a given part of the spectrum dies away in amount very rapidly when we pass from the maximum in the direction of decreasing wave-length; it dies away much more slowly as we pass in the opposite direction along the spectrum.

Cauchy's formula (§ 183) connecting the refractive index of a substance for a given radiation with the wave-length of that radiation does not agree with Langley's observations on the heat-rays. Briot's formula agrees better, but it ultimately differs from them in the opposite direction to that in which Cauchy's differs from them.

233. *Radiation at Different Temperatures.*—Hitherto we have assumed that the radiating bodies with which we have dealt have been kept at constant temperature so that their radiating powers remained unaltered. We must now consider the relation between emissivity and temperature.

Let us suppose that the hot body is placed inside an enclosure which is kept at a constant temperature t. Let $t+\theta$ be the temperature of the hot body, and let no heat pass from it except by radiation.

If r_t represent the rate of loss of heat from the hot body at temperature t, the rate of loss of heat under the assumed conditions will be

$$r_{t+\theta}-r_t;$$

and, since we have no adequate knowledge of the mechanism of emission, the form of this expression must be determined from experiment.

According to Newton, the rate of loss of heat is proportional to the excess of the temperature of the hot body over that of its surroundings; or, in symbols,

$$r_{t+\theta}=r_t+a\theta$$

But, since we must regard the rate of emission as independent of the surrounding bodies, this necessitates

$$r_t=at+b,$$

where a and b are constants, and t is *any* temperature.

The above law is not even roughly accurate unless the differences of temperature are small, and it becomes less and less applicable the greater the differences are.

Dulong and Petit made an elaborate series of experiments with the view of discovering a more correct law. They found that, when the temperature difference was kept constant, the rate of loss increased in geometrical progression as the temperature of the surroundings of the body increased in arithmetical progression; and the ratio of the geometrical series is independent of the temperature excess—so long, at least, as the excess is not greater than 200° C.

When the temperature excess vanishes, the loss of heat is zero. Dulong and Petit therefore expressed the law in the form

$$r_{t+\theta}-r_t=ab^t\,(b^\theta-1),$$

which agrees very closely with the result of their observations when t varies from 0° up to 80° C., and θ does not exceed 200° C. From this expression we see that the absolute rate of radiation, independently of the surroundings, is

$$r_t=ab^t+c,$$

where c is a constant.

The quantity a depends on the nature of the radiating surface, but the quantity b is practically an absolute constant, and is equal to 1·0077 if t be expressed in absolute units. The law of Dulong and Petit, therefore, asserts that, when the temperature excess is constant, the rate of loss of heat is proportional to $(1·0077)^t$, where t is the absolute temperature of the bodies to which the heat is radiated. It asserts also that the rate of loss is proportional to $(1·0077)^\theta - 1$ when t is constant.

The condition

$${}_1r_3 = {}_1r_2 + {}_2r_3,$$

where r represents rate of loss, might not have been borne out by experiment. The fact that it is so borne out furnishes, as Balfour Stewart pointed out, an independent proof of the truth of Prevost's Theory of Exchanges, of which it is a necessary consequence.

More recent experiments by De la Provostaye and Desains verified the accuracy of the law of Dulong and Petit within the limits already assigned.

The measurements of Hopkins give the following values for the constant a in the formula of Dulong and Petit, the unit of heat being the caloric (§ 152), and the emission being that from a square centimetre per second.

Glass	0·0121
Dry chalk	0·0100
Sandstone	0·0103
Polished limestone	0·0106
Unpolished limestone	0·0149

The factor 9·43 reduces these numbers to the value of the radiation when the temperature of the substance is 100° C., and that of the enclosure is 0° C. The same block of limestone was used in the two conditions.

Macfarlane gives the formula $0·000238 \times 3·06(10)\ {}^{6}t - 2\text{-}6(10)^{-8}t^2$ as expressing, in centimetre gramme second units, the emissivity of blackened copper at the temperature t°C. He gives the formula $0·000168 + 1·98(10)^{-6}t - 1·7(10)^{-8}t^2$ for polished copper.

Dulong and Petit's law seems to be applicable only to the total radiation from a body, and not to each definite radiation of which the whole is composed. The rate of emission of particular radiations from a black body seems to increase rapidly at first, and then more slowly, as the temperature of the black body is raised.

Using the observations of De la Provostaye and Desains, Stefan has shown that the results can be equally well explained on the assumption that the absolute rate of radiation is proportional to the fourth power of the absolute temperature. And Boltzmann and Galitzine have shown that this law results from the laws of thermo-dynamics and the electromagnetic theory of light. On the other hand, Bottomley considers that his recent experiments on emissivity, and those of Schleiermacher, do not bear out Stefan's law.

234. *Solar Radiation.*—Pouillet was the first to make fairly accurate measurements of the amount of radiation which is received by the earth from the sun in a given time. For this purpose he invented the *Pyrheliometer.*

This instrument consists of a flat cylindrical metallic vessel, the surface of which, with the exception of one end which is covered with lampblack, is highly polished. The bulb of a thermometer is inserted in the cylinder, and its stem lies along the axis. The cylinder is filled with water or mercury, and its blackened face is directed towards the sun. In order that this may be done with accuracy, a metal disc, the diameter of which is exactly equal to that of the cylinder, is fixed on the end of the axis of the instrument remote from the cylinder. The shadows of the disc and the cylinder will not coincide unless the face of the latter be accurately turned towards the sun. The area of the blackened face, and the amount of heat necessary to raise the temperature of the cylinder and its contents to a given extent, are exactly determined.

If the temperature of the instrument is the same as that of the air, and if its blackened face, carefully shaded from the sun, be turned towards the sky, radiation will take place, and the temperature will fall θ° (say) in t units of time.

Now let the apparatus be turned towards the sun for t units of time; after which let it be turned towards the sky as before, during an equal time, and let θ' be the fall of temperature.

It is concluded that the deficiency introduced into the total rise of temperature, which took place when the instrument was exposed to the sun, is

$$\frac{\theta+\theta'}{2};$$

for, during this exposure, the cylinder was steadily rising in temperature because of the heat which it absorbed from the sun, and was at the same time steadily falling in temperature because of its own radiation.

Hence the full rise of temperature, had no radiation taken place from the cylinder, would have been

$$\Theta+\frac{\theta+\theta'}{2},$$

where Θ is the rise of temperature which was actually observed. And so, from the known constants of the instrument, the quantity of heat which was received in a given time by unit area at the earth's surface could be calculated.

Pouillet gave the expression

$$ae^l$$

for the rate of rise of temperature under the above conditions at different times of the day. The quantity l is the thickness of the earth's atmosphere which was traversed by the sun's rays. The quantity e varies from day to day, but the quantity a does not so vary. Pouillet concluded that the expression would be applicable even if there were no atmosphere, in which case the constant a would represent the rate of rise of temperature; and he calculated that, in the event of no atmospheric absorption, the quantity of heat falling on a square centimetre of the earth's surface in one minute would raise the temperature of 1·76 grammes of water by 1° C.

From the average values of the quantity e, Pouillet inferred that about one-half of the total incident solar radiation is absorbed in the earth's atmosphere. Lord Kelvin concludes, from the above data, and those independently given by Herschel, that the rate of radiation from the sun's surface is about 7,000 horse-power per square foot, or thirty times the amount which is radiated per square foot from the furnace of a locomotive.

The *Actinometer* is another instrument used to determine the magnitude of solar radiation. It consists essentially of a metal enclosure (blackened internally and kept at constant temperature) at the centre of which the bulb of a thermometer is placed. By means of an opening in the enclosure, the sun's rays are allowed to fall upon the bulb for a given time, after which the bulb radiates to the enclosure. Various experimenters have used this apparatus in one or other of its forms. By its means Violle has found that the radiation from the sun, which would fall per minute on a square centimetre of the earth's surface did no absorption occur, would raise the temperature of 2·54 grammes of water by 1° C.

Langley's more recent measurements also indicate that Pouillet's estimate is too low. He finds that the quantity of unabsorbed heat falling on a square centimetre would raise the temperature of 1·81 grammes of water by 1° C. Did no absorption take place, this would become 2·8 grammes, and the amount of heat radiated from the sun per unit area of its surface would be fifty times greater than that radiated from a unit area of the furnace of a locomotive.

From Langley's data we can calculate that the yearly radiation from the sun to the earth would, if spread uniformly over the surface, melt a uniform crust of ice about 46 metres in thickness.

235. *Radiation from Moving Bodies.*—The following considerations regarding the motion of radiating bodies, first advanced by Balfour Stewart, are specially worthy of notice—apart from their intrinsic value—as an example of the useful employment of the principle of conservation of energy.

Let us suppose that ultimate equality of temperature has been arrived at inside the enclosure, impervious to heat, which was postulated in § 177; and let us further assume that one of the bodies in that enclosure suddenly commences to move about with a speed which is comparable with that of light. A direct application of Döppler's principle shows us that any other body in the enclosure, towards which the motion may be directed, will receive energy from the moving body at a greater rate than will one which is so situated that the motion is directed from it.

It follows that relative motion of radiating bodies is inconsistent with ultimate equality of temperature. But persistent inequality of temperature would imply a perpetual source of energy; and we therefore conclude that the relative motion of radiating bodies must gradually cease.

CHAPTER XX.

EFFECTS OF THE ABSORPTION OF HEAT: DILATATION.

236. *Temperature.*—Increase of temperature usually accompanies the application of heat to any substance. The fundamental distinction between *heat* and *temperature* is very clearly brought out by a consideration of the meaning of the words *hot* and *cold* as applied to different material substances. The bodies which are said to be hot or cold may really be at the same temperature. Thus a mass of iron and a mass of wood, though their temperatures be equal, *feel* very differently to the touch. If the temperature of the hand be higher than that of the two masses, the former feels cold, and the latter feels warm; while, if the temperature of the hand be lower than that of the bodies, these conditions are reversed. The reason is that the physical properties of the substances are such that, of the two, iron is the one which most rapidly extracts heat from, or supplies heat to, the hand.

It is sufficient for our present purpose that we regard temperature as that condition which determines the flow of heat from one body to another (see §§ 228, 150).

If two bodies, which differ in temperature, be placed in contact with each other (or in thermal communication of any sort), heat passes from the body which is at the higher temperature to the body at the lower temperature. And two bodies, between which, *on the whole*, there is no transference of heat when thermal communication is established, are said to have *equal* temperatures.

It is an experimental fact that any two bodies, which have each the same temperature as a third body, are themselves at equal temperatures.

Various methods, which will be described later, are used for the determination of the difference of temperature which subsists between two bodies, or between a body in one given physical state and the same body when in another state.

We may premise, for our present purpose, that the temperatures of ice-cold water, and of boiling water, respectively, are called 0 degrees and 100 degrees; that a given number of degrees, at any part of our scale, corresponds to a constant interval of temperature; and that we call the degrees *Centigrade* degrees—the letter *C* being used to discriminate them from the degrees of other scales.

237. *Dilatation of Solids.*—One of the most obvious effects of the application of heat to a substance is expansion. In a homogeneous isotropic solid, the expansion is equal in all directions. On the other hand, in a non-isotropic solid, the expansion is different in different directions; but, in this case, three rectangular directions (called the principal axes, *cf.* § 219) can always be found such that in one the expansion is a maximum; in another, it is a minimum; and, in the third, its value is intermediate between those along the other two—being, in fact, a maximum-minimum. When the expansions, along these directions, which accompany a given rise of temperature, are known, the expansion in any other direction can be found.

We have, first of all, then, to consider the laws of linear dilatation of a solid. [The requisite measurements are readily made by direct micrometric methods, which give accurate determinations of the lengths of a bar at different known temperatures.]

1. It is found that *the increase of length of a given bar is proportional to the length of the bar at the initial temperature.*

2. *The alteration of length is proportional to the increase of temperature.*

These laws are symbolically represented by the equation

$$l_t = l_0(1+kt),$$

where l_t and l_0 are respectively the lengths of the bar at the higher and lower of the two temperatures of which t is the difference, and k is a constant called the *co-efficient of linear dilatation.* This constant is obviously the increase of length, per unit rise of temperature, of a bar the original length of which is unity.

From the three equations

$$l_t = l_0(1+k_1t).$$
$$b_t = b_0(1+k_2t),$$
$$d_t = d_0(1+k_3t),$$

where the letters l, b, and d denote the length, breadth, and thickness of a rectangular parallelepiped of the substance, and k_1, k_2, k_3, are respectively the co-efficients of dilatation measured in the direction of the length, breadth, and thickness, we can obtain an expression for the cubical dilatation of the solid. Let us suppose that these equations refer to the principal axes; then

$$l_t b_t d_t = l_0 b_0 d_0 (1+k_1 t)(1+k_2 t)(1+k_3 t),$$

or

$$V_t = V_0 (1+k_1 t)(1+k_2 t)(1+k_3 t),$$

where the Vs represent volumes.

In all cases in nature, the quantities k_1, k_2, k_3 are so small that their squares and products may be neglected; hence, to a sufficient degree of approximation,

$$\begin{aligned} V_t &= V_0 (1+(k_1+k_2+k_3)t) \\ &= V_0 (1+Kt), \end{aligned}$$

if K represents the *co-efficient of cubical dilatation*. This gives

$$K = k_1 + k_2 + k_3;$$

and so the co-efficient of cubical dilatation (which is the fractional increase in bulk of unit volume per unit rise of temperature) is equal to the sum of the three principal co-efficients of dilatation. When the substance is isotropic, this latter equation becomes

$$K = 3k;$$

which asserts that the co-efficient of cubical dilatation of a homogeneous isotropic solid is three times the co-efficient of linear dilatation.

If the edges of the rectangular parallelepiped be not in the directions of the three principal axes, the effect of the application of heat will be to change the inclination of the faces, so that the parallelepiped will cease to be rectangular. When the three edges are all originally equal, and the diagonals of one of the square faces of the cube are parallel to two of the principal axes, the application of heat changes the square face into a rhombus such that the tangent of half its obtuse angle is equal to $1+(k_1-k_2)t$, where t is the increase of temperature and the original length of the diagonals is assumed to be unity. This affords a ready means of determining the co-efficient of expansion along one of the two principal axes, provided that we know its value along the other; for the change of angle can be measured with extreme accuracy by optical methods.

Fizeau introduced a specially delicate method of measuring the change of length of a rod. Newton's rings (§ 195) are produced by the interference of light reflected at the surface of a film of air contained between two plates of glass, of which one at least is slightly curved; and the colour of any particular ring depends on the thickness of the film, which can be calculated when the radius of the ring is known. If one of the two plates be fixed, while the other is attached to the expanding rod, the slightest change of length of the rod causes a diminution of the thickness of the film of air which can be easily calculated by means of the optical changes which are simultaneously produced.

He gives the following empirical formula connecting the co-efficient k with the temperature t expressed in Centigrade degrees:

$$k=a+a'(t-40).$$

The constants were determined by observations at three temperatures, viz., 10°, 45°, and 70°.

The cubical dilatation of a solid is usually determined directly by heating it in a liquid, of known expansibility, which is contained in a vessel made of a substance whose co-efficient of linear (and, therefore, also of cubical) dilatation has been found. Let V_o be the volume of the vessel at 0° C., and let K_1 be its co-efficient of dilatation. The volume at t° C. is

$$V_o(1+K_1t).$$

Similarly, the volume at t° C. of the liquid which filled the vessel at 0° C. is

$$V_o(1+K_2t).$$

If we now place in the vessel a solid, the volume of which at 0° C. is v_o, and at t° C. is

$$v_o(1+K_3t),$$

the volume of liquid which overflows from the vessel at temperature t is

$$V_o(1+K_2t)-V_o(1+K_1t)+v_o(1+K_3t),$$

or

$$V_o(K_2-K_1)t+v_o(1+K_3t),$$

from which expression the value of K_3 may be found.

It is needless to enumerate the various practical applications of the dilatation of solid bodies when their temperature is raised, or of their contraction as the temperature falls. The well-known processes of shrinking the tires on wheels, and of drawing together the walls of a building when they have bulged outwards, will sufficiently serve as an indication of their nature.

One particular application—to the construction of a compensation-pendulum or balance-wheel of a watch—merits special notice.

The ordinary compensation-pendulum is constructed upon the principle that the *difference* between the lengths of two rods, of different expansibilities, will remain constant, however the temperature may be altered within allowable limits, provided that the lengths of the rods be made in inverse proportion to the co-efficients of dilatation of the substances of which they are composed. This obviously affords a means of keeping constant the distance between the bob of a pendulum and its point of support.

If two equal bars, of different expansibilities, be soldered together throughout their length, a rise of temperature will cause the compound bar to bend in such a way that the less expansible bar is on the concave side; for this is the only way in which the tendency towards unequal expansions can be satisfied. This fact is made use of in the construction of compensated balance-wheels. When the temperature of an uncompensated wheel increases, the expansion of the spokes carries the rim of the wheel further out from the centre, and the consequent increase of the moment of inertia produces an increase in the period of oscillation. Hence a watch, or chronometer, fitted with such a wheel will go too slow in warm weather and too fast in cold weather. But if the rim of the wheel be divided into a number of independent parts, each part being carried by a separate spoke, and if the rim be compound as above described, matters may be so arranged that the throwing out of the weight from the centre because of the expansion of the spokes is counterbalanced by means of the inward bending of the segments of the rim.

It is worthy of note that one, or two, of the principal expansibilities of a solid may be negative, *i.e.*, the substance may contract in at least one direction when its temperature is raised. In such a case a series of directions may be found in the substance such that a change of temperature does not give rise to any alteration of length of the substance in any of these directions. These directions may be found by imagining a sphere to be drawn in the unheated body and finding its intersection with the ellipsoidal surface into which it becomes distorted by the application of heat. All lines drawn in the body parallel to the lines joining the points of the curves of intersection to the centre of the sphere are unchanged in length. Hence a rod cut from the substance in any such direction may be used as the rod of a compensation pendulum. Brewster pointed out that a rod of marble might be so used.

An interesting example of contraction of a solid when its temperature is raised is seen in the case of indiarubber under tension. The experiment may readily be performed by blowing steam through a hollow tube of the substance, which is fixed at its upper end, and is extended by means of a weight attached to the lower end.

The subjoined table contains Fizeau's determinations (see above) of the values of the constants in his formula for the co-efficients of linear dilatation of a few well-known substances. The constant a is, of course, the value of the co-efficient at 40°. When more than one value is given these refer to the various principal dilatations.

	a.	a'.
Carbon (retort)	0·00000540	0·0000000144
Platinum	0·00000905	0·0000000106
Steel	0·00001095	0·0000000124
Iron (compressed)	0·00001188	0·0000000205
Copper (native)	0·00001678	0·0000000205
Silver	0·00001921	0·0000000147
Lead	0·00002924	0·0000000239
Quartz	0·00000781	0·0000000205
	0·00001419	0·0000000348
Iceland spar	0·00002621	0·0000000160
	0·00000540	0·0000000087
Aragonite	0·00003460	0·0000000337
	0·00001719	0·0000000368
	0·00001016	0·0000000064

An average value of the co-efficient of expansion of glass is 0·0000085.

238. *Dilatation of Liquids.*—In liquids it is merely the cubical dilatation with which we have to deal. This may readily be determined if we know the cubical dilatation, K$'$, of the substance of which the containing vessel is composed. Let K be the unknown co-efficient, and let an amount of the liquid, of weight W, fill the vessel (which must have a narrow neck) at a temperature t, while the weight of the quantity which fills the vessel at 0° is W_o. Now, the weight of the quantity which would have filled it, had the liquid been inexpansible, is $W_o(1+K't)$; but, since the liquid is expansible, this amount is diminished in the ratio of unity to $1+Kt$. Hence we get

$$W(1+Kt)=W_o(1+K't),$$

which determines K.

A very simple method, by means of which the co-efficient of dilatation of a liquid may be found without any knowledge of that of the substance of which the containing vessel is composed, was devised by Dulong and Petit. In all essential particulars the apparatus consists of a double U-tube, *abcd* (Fig. 132). The portion *bc* is occupied by air, which separates the portions of the liquid contained in *ab* and *cd* while preserving continuity of pressure between them. Equal atmospheric pressure acts at the points *a* and *d*, and the air in *bc* is necessarily at uniform pressure throughout. Hence, when equilibrium is maintained, the pressure per square inch due to the difference of level ab must be equal to the pressure per square inch due to the difference of level dc. Let us suppose that the limb cd is raised to temperature t, while the limb ab is kept at 0°. The average expansibility throughout the range of temperature is

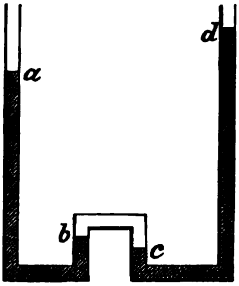


FIG. 132.

$$\frac{dc-ab}{ab\,.\,t}.$$

If v be the increase of volume produced by a small rise of temperature t, while V_0 is the total volume of the liquid at 0°, the

quantity $$\frac{v}{V_0 t}$$

is usually taken as the co-efficient of dilatation at the given temperature. The *true* co-efficient at any temperature is obviously given

by the ratio $$\frac{v}{Vt}$$

where V is the volume of the liquid at that temperature.

The values, given by Regnault, of the ordinary co-efficient of dilatation of mercury at temperatures varying from 0° C. to 350° C. are well represented by the formula

$$K=0{\cdot}0001791+0{\cdot}0000000504t.$$

Water presents a marked peculiarity as regards its change of volume with rise of temperature. Between its freezing-point, 0° C., and a temperature of almost exactly 4° C., its volume diminishes as the temperature increases. At all higher temperatures the volume increases as the temperature is farther raised. Thus water is in a condition of *maximum density* at a temperature of about 4° C.

The existence of the temperature of maximum density is readily shown by means of *Hope's experiment.* The necessary apparatus consists of a cylindrical glass vessel the central portion of which is surrounded by a metal casing in which a freezing mixture may be placed. Two thermometers are inserted horizontally through the glass vessel, one near the top and the other near the bottom, so that their bulbs are at the axis of the cylinder. The vessel being filled with water, a freezing mixture is placed in the casing. Very soon the temperature marked by the lower thermometer begins to diminish, which shows that the cold water is of greater density than the warm water near the top of the vessel. This goes on until the lower thermometer registers a temperature of 4° C., at which it remains. Soon afterwards the temperature of the water at the top of the vessel begins to fall, which shows that the colder water is now ascending and must therefore be expanding; and this process goes on until the water at the top freezes at 0° C.

The maximum-density point is lowered by pressure to the extent of about 3° C. by a pressure of one ton's weight per square inch.

Kopp's determinations of the co-efficient of dilatation of water are fairly well represented between 0° C. and 20° C. by the formula

$$K = \frac{t - 4}{72{,}000}.$$

The more recent experiments of Pierre, Hagen, and Mathiessen indicate that the denominator of this fraction is about 5½ per cent. too large; but Rossetti's results agree better with those of Kopp.

Different experimenters have determined the co-efficients of expansion of various liquids when under pressure sufficient to keep the liquid in equilibrium with its vapour at temperatures above the ordinary boiling-point. Drion gives the subjoined values of the co-efficient of dilatation of sulphurous acid:

Temp. C.	Co-efficient.	Temp. C.	Co-efficient.
0°	0·00173	70°	0·00318
10°	0·00188	90°	0·00415
30°	0·00219	110°	0·00592
50°	0·00259	130°	0·00957

From these results it appears that the co-efficient of dilatation of this liquid, at about 120° C., is double of that of air.

Hirn gives the volume of water at different temperatures as follows:

Temp. C.	Volume.	Temp. C.	Volume.
4°	1·00000	140°	1·07949
100°	1·04315	160°	1·10149
120°	1·05992	180°	1·12678

Consequently, at 180° C., the expansibility of water is nearly one half of that of air.

239. *Dilatation of Gases.*—A gas must be kept in an enclosed space when experiments are to be made upon it with regard to its alterations of volume under varying conditions of temperature. But we know that, so long as the temperature is maintained at a constant value, the volume and pressure of the gas are in inverse proportion to each other. Hence we may investigate the effect of variation of temperature in two ways: we may measure the expansion under constant pressure; or, we may measure the change of pressure at constant volume.

By such measurements Charles (and, subsequently, Gay Lussac) was led to the conclusion that *the volume of a given quantity of any gas, under constant pressure, increases by a constant fraction of its amount for a given increment of temperature.* This statement is known as *Charles' Law*, and is represented symbolically in conjunction with Boyle's Law by the equation

$$pv = C(1+at),$$

where c and a are constants, and the meaning of the other quantities is obvious.

The pressure being kept constant, the volume increases by the fraction a of its amount at 0° per unit rise of temperature; a is therefore the co-efficient of dilatation under constant pressure. Again, if the volume be kept constant, the pressure increases by the fraction a of its amount at 0° per unit rise of temperature. Thus the fractional increase of volume under constant pressure, and the fractional increase of pressure at constant volume, have the same numerical value if the above equation be rigidly true.

The magnificent series of experiments carried out by Regnault have shown that, while the above law is very nearly true in the cases of the most permanent gases, marked discrepancies are exhibited when the more readily liquefiable gases are employed. The experiments on the undernoted gases gave as the value of the dilatation between 0° C. and 100° C.:

Hydrogen	0·3667	0·3661
Air	0·3665	0·3670
Nitrogen	0·3668	0·3670
Carbonic oxide	0·3667	0·3669
Carbonic acid	0·3688	0·3710
Cyanogen	0·3829	0·3877
Sulphurous acid	0·3845	0·3903

The figures in the second column represent the value of the co·efficient as determined at constant volume; those in the third column are the values determined under constant (atmospheric) pressure. With a single exception (in the case of hydrogen) the latter number exceeds the former.

When the initial pressure of gases at 0° is increased, the dilatation between 0° and 100° increases. Some of Regnault's results for air are:

109·72	149·31	0·3648
374·67	510·35	0·3659
760·00	1038·54	0·3665
1692·53	2306·23	0·3680
3655·66	4992·09	0·3709

The numbers in the first two columns represent respectively the pressure per unit area at 0° and at 100° expressed in terms of the weight at 0° of a column of mercury of unit section and one millimetre in height.

For carbonic acid Regnault gives the similar results:

758·5	1034·5	0·36856
901·1	1230·4	0·36943
1742·9	2387·7	0·37523
3589·1	4759·0	0·38598

The variation in the case of this gas is therefore greater than in the case of air.

The results, for the same two gases, under constant pressure, are:

Air	760	0·36706
	2525	0·36944
Carbonic acid	760	0·37099
	2520	0·38455

Regnault's apparatus consisted essentially of a glass bulb D (FIG. 133) which contained the gas and communicated through the tube E with a reservoir AB, which was filled with mercury. A tube F, open to the atmosphere, also communicated with the reservoir. When the co-efficient was to be determined at constant volume, a plug C was screwed in until the mercury stood at the points E and F in the tubes—E being a fixed point. D was, under these conditions, surrounded successively by melting ice, and by the steam from boiling water, and the pressure in each case was found from the difference of level of the mercury in the two tubes and the known barometric pressure. Suitable corrections had, of course, to be applied for the expansion of the bulb by heating or by pressure.

When the dilatation was observed under constant pressure, the plug C was, at each temperature, screwed out until the mercury stood at the same level in the two tubes. The pressure was then equal to that of the atmosphere.

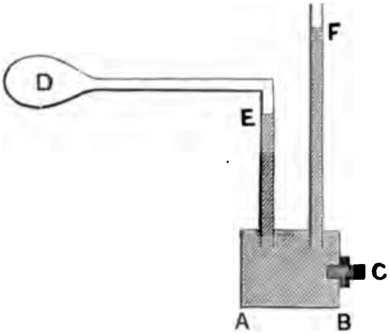


FIG. 133.

Other volumetric, or gravimetric, methods have been employed by Regnault and various experimenters.

240. *Absolute Zero of Temperature.*—We may write the equation which expresses Boyle's and Charles' Laws in the form

$$pv = C\alpha\left(\frac{1}{\alpha} + t\right).$$

Now as the second term within the brackets represents temperature, the first term must also represent a temperature, for the dimensions (§ 22) of all the terms of a physical equation must be identical. And the numerical value of that temperature is about 273, since $\alpha = 0{\cdot}003665$. Hence we must suppose that the zero of the Centigrade scale corresponds to a temperature of 273 degrees on this new scale, which we may call a scale of *absolute* temperature, since its magnitude is independent of the particular gas employed.

To look at the matter from another point of view, we observe that, as t diminishes to zero and then becomes an increasing negative quantity, the product pv constantly diminishes, and finally becomes zero when $t = -273°$. If the volume is constant, this means that the pressure vanishes when t has this value. But the pressure of a gas is due to the motion of its particles, and hence, when $t = -273°$ the particles cease to move, and the gas is therefore totally deprived of heat.

The above equation may therefore be written in the form

$$pv = \mathrm{R}t,$$

where $\mathrm{R} = \mathrm{C}a$ and t represents absolute temperature.

We shall see later (§ 303) that this estimate of the position of the absolute zero on the Centigrade scale is confirmed by thermodynamical considerations.

241. *Measurement of Temperature.*—The most usual method of measuring temperature is by means of the expansion of a liquid or a gas. Mercury is generally used in the former case, air in the latter.

A glass tube of narrow, and as nearly as possible uniform, bore is first chosen. If the bore be not quite uniform, its variations are determined by the process of calibration. This consists in running a small quantity of mercury along the tube from one end to the other, and measuring its length at the various parts. The quotient of the weight of the mercury by the product of its specific gravity and the length of the column at any part gives the mean section of that portion of the tube. Next, a bulb is blown, on one end of the tube, of a size which is determined by the bore of the stem, the expansibility of the liquid to be used, and the required length of a scale division. The bulb is now slightly heated to expel some air, and the instrument is then inverted in a vessel of the liquid with which it is to be filled. As the bulb cools, some of the liquid enters it. This liquid is then boiled in the bulb, and its vapour expels the remaining air. A repetition of the process of inversion of the bulb and stem in the liquid will result in both being entirely filled. So much of the liquid is then run out that the remainder scarcely fills the stem when it is boiled once more. The vapour drives out the air which entered, and the tube is then hermetically sealed.

Two fixed points must now be determined on the stem. As we shall see afterwards (§§ 247, 249), this may be done by means of melting ice, and of the steam coming from boiling water in a nearly closed vessel.

But, before these points are determined, a considerable time should be allowed to elapse, for the bulb will not shrink quickly to its final volume. The process of shrinkage usually goes on for years, though, by careful annealing, the effect may be considerably lessened.

To determine the lower fixed point of the scale, the bulb and part of the stem are surrounded by melting ice. The final position of the extremity of the liquid column is marked on the stem.

The upper fixed point is determined by surrounding the bulb, and as much of the stem as possible, by the steam which is escaping from water boiling under a pressure of one atmosphere. The final position of the liquid is then marked on the stem.

The distance between the two marked points is then divided into a number of equal parts.

On the Centigrade scale, the lower fixed point is marked 0°, and the higher is marked 100°; on the Fahrenheit scale the lower is marked 32°, and the higher 212°. Thus, on the Centigrade scale, there are 100 divisions between the boiling-point and the freezing-point of water—hence its name; on the Fahrenheit scale there are 180 divisions between these points. The Fahrenheit zero was determined by a freezing-mixture of snow and salt, which gave the lowest temperature known at the time when Fahrenheit first constructed his thermometers. The relation between the two scales is obviously given by the equation

$$\frac{F-32}{180}=\frac{C}{100},$$

where F and C respectively represent the Fahrenheit and Centigrade scale readings.

The same interval on Réaumur's scale is divided into 80 equal parts, the zero being the same as that of the Centigrade scale.

As already remarked, the liquid generally used is mercury. For the measurement of temperatures below the freezing-point of mercury, alcohol is employed.

The air thermometer is of great use in the determination of temperatures above those at which mercury can be employed; and its readings agree pretty closely with those of the true absolute scale of temperature as determined by thermodynamical considerations. The following numbers, taken from Regnault's results, show the difference between the Centigrade scale of the air thermometer and that of the mercury thermometer:

Air ...	... 0°	20°	40°	60°	80°	100°	200°	300°
Mercury	... 0°	19°·98	39°·67	59°·62	79°·78	100°	202°·78	308°·34.

Another point which tells greatly in favour of this thermometer is the large expansibility of air. Because of this, the instrument is much more free from errors due to the irregular expansion or contraction of the glass than are liquid thermometers. The air which is used must be carefully freed from water vapour and impurities such as carbonic acid.

There are two ways in which the instrument may be employed. The air may be kept under constant pressure while the variation of ts volume is noted, or it may be maintained at constant volume while the change of pressure is noted.

In the constant-pressure air thermometer great difficulty arises from the correction which has to be applied for the air which fills the stem of the instrument. So serious is this, that Regnault found that it was practically impossible to get good results from a constant-pressure gas thermometer. The essential features of his thermometers have been already described in § 239. Callender has recently devised an instrument in which the difficulties are overcome. The bulb which contains the air is connected with another bulb which contains mercury, the amount of which can be varied at pleasure. A third bulb contains air, which is kept at 0°C. by means of melting ice. To this bulb is attached a tube, with a closed end, exactly similar to the tube which connects the air bulb and the mercury bulb. These similar tubes are placed so near together that they are always at a common temperature, and they are connected, near the middle of their length, by means of a U-tube partly filled with sulphuric acid. When the instrument is in use, the amount of mercury in the one bulb is varied until the sulphuric acid stands at the same level in both limbs of the U-tube. The pressure of the air in the bulb which is used to measure the temperature is then equal to the known pressure of the air in the bulb whose temperature is kept at 0° C. If the mass of the air in the latter bulb is equal to that of the air in the other air bulb and the mercury bulb, no correction has to be applied because of the air in the connecting tube. The mercury bulb may also conveniently be kept in melting ice.

In *the vapour-pressure thermometer* observation is made of the pressure of a vapour in contact with its liquid, which (§ 249) is determined by the temperature.

Self-registering thermometers are frequently employed for the purpose of indicating the maximum, or the minimum, temperature attained between two given periods. In the usual form of the maximum thermometer, the expanding column of mercury pushes an iron index along the tube. This index is left behind when the column contracts, for mercury does not wet iron. In the usual minimum thermometer, alcohol is used along with a glass index. The liquid, when expanding, flows past the index ; when it contracts, it pulls the index with it, for its surface tends to take the smallest possible area (§ 96), and this is attained when it occupies the space between the index and the walls of the tube.

Continuously-registering thermometers are also used. The principle of one of the best forms of these instruments is identical with that of the Bourdon gauge. Let *abcd* (Fig. 134) be a longitudinal section of a hollow metal receiver in its unstrained state, and let us suppose that the receiver is filled with a liquid. As the temperature rises, the pressure increases; and, since *cd* is greater than *ab*, the sum of the moments of the forces tending to straighten the receiver exceeds the sum of those which act contrariwise. A system of levers attached to such a receiver (fixed at one end) traces a continuous record on a properly-graduated paper placed on a drum which revolves slowly at a uniform rate.

Pyrometers are used for the determination of very high temperatures. In Daniell's pyrometer, a bar of platinum is slipped into the hole bored in a rod of graphite. A plug of graphite, or baked clay, rests on the top of the bar and fits tightly into the hole, or is otherwise kept tightly in position. When the platinum (which rests on the bottom of the hole) expands, the plug is pushed out, and remains in its position of maximum displacement when the temperature falls. From the increase of length of the platinum thus registered, we can calculate the temperature of any furnace in which it may be placed, *on the assumption* that the law of dilatation, determined throughout moderate ranges of temperature, holds up to the high temperatures of the furnace.

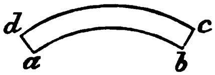


FIG. 134.

Another method (first used by Regnault) consists in boiling mercury in a flask of wrought iron until the air is expelled by the vapour and no liquid is left. When equilibrium of temperature is attained, the flask is closed and allowed to cool, so that the vapour condenses. In this way the density of the vapour at the high temperature can be found, and from this the temperature can be calculated on the assumption that the vapour obeyed Boyle's Law. Deville and Troost employed the vapour of iodine in a porcelain flask.

Siemens introduced the method of determining high temperatures by means of measurements of the electric resistance of a platinum wire (Chap. XXVI.). This method is found to be capable of great accuracy.

In another method thermo-electric circuits are used (Chap. XXV.).

The method of calorimetry, in which the amount of heat given out by a known mass of (say) platinum, whose specific heat is known, is measured, is not capable of great accuracy.

CHAPTER XXI.

EFFECTS OF THE ABSORPTION OF HEAT: CHANGE OF TEMPERATURE AND CHANGE OF STATE.

242. *Unit of Heat. Specific Heat. Thermal Capacity.*—One of the most marked effects of absorption of heat is a rise of the temperature of the heated body. In some cases no change of temperature takes place, and the effect which appears instead is a change of the physical state of the substance. But, before discussing these effects, we must consider the methods of measuring the amount of heat required to produce a given change; and this, in turn, necessitates the adoption of a definite unit of heat.

We may conveniently adopt as our unit *the quantity of heat which is required to raise the temperature of one gramme of ice-cold water to* 1° C. This unit is called the *calorie.*

A given quantity of heat may therefore be measured by the number of grammes of ice-cold water which it can raise in temperature to 1° C. We might measure it also by the number of grammes of ice at 0° C. which it is just able to melt; for, as we shall see shortly, the number of units of heat which are required to just melt one gramme of ice at 0° C. is quite definite and measurable. More generally, we might adopt as our unit the amount of heat necessary to produce any definite physical change, and then we might determine what fraction of the unknown quantity of heat could produce this change. The number of units in the unknown quantity would be the reciprocal of this fraction.

We define the specific heat of a substance, under given conditions, *as the quantity of heat which is required to raise the temperature of one gramme of the substance by* 1° C. From this definition, and from our previous definition of the unit of heat, it follows that we must consider the specific heat of one gramme of water at 0° C. to be unity.

More strictly, we should define it as the rate at which the quantity of heat supplied to the substance, per gramme of its mass, varies with the temperature. But, in all actual cases, there is no practical difference between the two definitions.

The *mean*, or *average*, specific heat of a substance, throughout a given range of temperature, is obtained by dividing the amount of heat which is required to raise one gramme of the substance through the given range by the difference between the two extreme temperatures.

The *Thermal Capacity* of a substance is the quantity of heat which is required to raise the temperature of unit volume of the substance by one degree. It is therefore equal to the product of the specific heat and the density of the substance.

243. *Specific Heat of Solids and Liquids.*—Various methods are used for the determination of specific heat.

In one method, use is made of the fact that the rate of emission of heat from a body at a given temperature depends only upon the nature of its surface. Hence, if we fill a thin metal globe successively with two different liquids, and observe the rate of cooling of each liquid at the same temperature, and under the same conditions otherwise, we can compare the specific heats of the two liquids. For, if m, s, r, and m', s', r', represent respectively the mass, the specific heat, and the rate of cooling of the liquids, we have

$$msr = m's'r'.$$

If one of the liquids be water, so that $s=1$, we get

$$s' = \frac{mr}{m'r'}.$$

In an actual experiment, the liquids would be raised to a common high temperature, and readings of their temperatures would be taken at equal intervals of time as they cooled. If a curve were then drawn the ordinates of which represented temperature, while the abscissæ represented time, the rate of cooling would be given by means of the tangent to the curve. Thus (Fig. 135), to find the rate of cooling at a temperature θ, draw the tangent ab to the curve at the point P, corresponding to θ, and let it intersect the time-axis in the point a and the temperature-axis at the point b; the rate of cooling is ob/oa.

Another method of determining specific heat is known as the *Method of Mixture.* Let m pounds of one substance at temperature t be mixed with m' pounds of another at temperature t', and let the specific heats of the two substances be respectively s and s'. If the temperature of the mixture be θ, the heat lost by the hotter body (say that of mass m') is $m's'(t'-\theta)$. Similarly, the heat gained by the colder substance is $ms(\theta-t)$. Also, if μ be the mass of the vessel which contains the substances (and we may regard the stirring-rod used to mix the substances as forming part of it) while σ represents its specific heat, the heat given to the vessel is $\mu\sigma(\theta-t)$. Here we assume that the colder substance was originally contained in the vessel, the hotter substance being introduced from without.

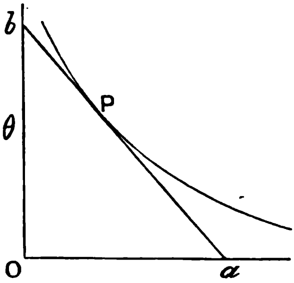


FIG. 135.

Instead of $\mu\sigma$ we may write w—a unit multiplier being understood. The meaning of this is that $w=\mu\sigma$ is the number of pounds of water (specific heat = unity) which require the same supply of heat to produce the rise of temperature $(\theta-t)$ as the vessel required. The quantity $\mu\sigma$ is therefore called the *water-equivalent* of the vessel.

We may therefore write

$$m's'(t'-\theta)=(m+w)(\theta-t),$$

provided that the cold substance is water. If the value of w be known, this equation enables us to find the value of s'. If w be not known, a second experiment in which the mass, m, of water is varied, will lead to another similar equation by means of which w may be eliminated or determined. Of course, the value of w may be found by one experiment alone, in which two quantities of water, at different temperature, are mixed.

In an accurate experiment of this kind, precautions are taken that there shall be as little loss of heat by radiation as possible. This may be effected by making the vessel which contains the water (or other liquid) of a substance which is a bad radiator of heat; and, in addition, this vessel is placed inside a second similar vessel, contact between the two being prevented by means of bodies which do not readily conduct heat. If necessary, a third vessel may be used; and then a correction may be made for the slight amount of heat which is still lost by radiation.

A third method of determining specific heat is by the *Fusion of Ice.* As we shall see subsequently (§ 248), a definite amount of heat is required to just melt one gramme of ice. Let H be this quantity. If M pounds of a substance melt m pounds of ice in the process of cooling from T° C. to 0° C., the average specific heat, S, throughout that range of temperature is given by the equation

$$m\mathrm{H} = \mathrm{MST}.$$

Bunsen and others have used forms of apparatus in which the quantity m is found by means of the decrease of volume of a mixture of ice and water when the heat which is given out by the cooling body melts some of the ice.

The *Method of Evaporation* is strictly analogous, and is of considerable practical use.

The *Method of Condensation* has been successfully used by Joly, and is of special scientific importance, for by it the specific heats of solids, liquids, and gases can be found; and Joly has further shown that accurate estimates of the specific heats of gases at constan volume, otherwise so difficult of determination (§ 245), may be got. One pan of a balance is suspended in a chamber into which steam can be either suddenly or slowly admitted. In that pan is placed the substance whose specific heat is to be found. Steam condenses on the pan and the substance, and weights are put in the other pan to restore equilibrium. In this way the quantity of steam which it is necessary to condense in order to raise the temperature of the pan and the substance through a given range of temperature is found. The fraction of this which is required for the pan alone is got by a separate experiment. Obvious corrections have to be applied and precautions taken. By making both pans alike—thermally and otherwise—and enclosing them both in the steam chamber, a differential steam calorimeter is formed in which no correction for the pan is needed. It is this form of the instrument, with equal copper globes—one empty, the other filled with the substance under investigation—attached to the pans, which is used for the determination of the specific heat of a gas at constant volume.

In general the specific heat of a substance increases with rise of temperature. But the specific heat of platinum varies very slightly with temperature, so that the range of temperature through which a mass of that metal cools is closely proportional to the quantity of heat which it emits. As has been already mentioned, this fact is utilised in the measurement of high temperatures.

Regnault found that the quantity of heat required to raise the temperature of a gramme of water from 0° C. to t° C. is represented by the equation

$$H = t + 0·00002t^2 + 0·0000003t^3,$$

and that therefore the true specific heat, at any temperature, is given by

$$1 + 0·00004t + 0·0000009t^2.$$

His experiments were carried out at various temperatures between 0° C. and 230° C. His results in the cases of some other liquids are:

Alcohol	$0·548 + 0·00224t + 0·00000662t^2$	-23	$+66$
Turpentine... ...	$0·415 + 0·00124t - 0·00000398t^2$	0	150
Ether	$0·529 + 0·000592t$	-20	$+30$
Carbon Bisulphide	$0·235 + 0·000163t$	-30	$+40$
Chloroform ...	$0·232 + 0·000101t$	-30	$+60$

The numbers in the last two columns indicate the range of temperature over which the respective observations extended.

By very careful measurements, Rowland found that the specific heat of water has a minimum value at a temperature of about 30° C. The existence of this minimum has been subsequently verified.

The specific heat of ice is almost exactly equal to 0·5, and Regnault found that it is diminished by decrease of temperature.

This table gives the specific heat of various elementary substances at ordinary temperatures:

	Solid.	Liquid.
Water	0·500 (at 0° C.) ...	1·000
Glass	0·180	
Iron	0·114	
Copper	0·096	
Zinc	0·094	
Silver	0·057	
Tin	0·056	0·064
Mercury	0·031	0·033
Lead	0·031	0·040

It is worthy of notice that the specific heat of water is considerably in excess of that of any of the other substances; and that, in general, the specific heat of any substance in the liquid state exceeds that of the same substance in the solid state.

244. *Law of Dulong and Petit.*—Dulong and Petit found that the product of the specific heat of any elementary solid into its atomic weight is practically constant. An alternative statement is

that *the water-equivalent of an atom of each elementary solid is practically constant.* The numbers in the first column below represent specific heat; those in the second column represent atomic weight; and those in the third give the product of these two quantities.

Iron ...	0·114	54·5	6·2
Copper	0·096	63·5	6·1
Zinc ...	0·094	64·5	6·1
Silver ...	0·057	108·0	6·2
Tin ...	0·056	118·0	6·6
Mercury (liquid)	0·033	202·0	6·6
Lead ...	0·031	207·0	6·4

Similar results have been established for various series of chemical compounds. The value of the constant varies from one such series to another.

If there be n_1 atoms of a substance whose atomic mass is m_1, and so on, in a molecule of a compound, the resultant molecular mass will be

$$M = n_1m_1 + n_2m_2 + \dots\dots$$

If the respective specific heats of the substances in the compound be s_1, etc., the molecular specific heat of the compound will be given by S in the formula

$$MS = n_1m_1s_1 + n_2m_2s_2 + \dots\dots$$

Now the products m_1s_1, etc., represent the atomic water equivalents whose common value for elementary substances in the uncombined condition may be taken as 6·38. Hence the molecular water equivalent is

$$(n_1 + n_2 + \dots)6{\cdot}38,$$

provided that the common value is not altered by combination. Regnault has verified this result in the case of metallic alloys, the quantity of each constituent being made a multiple of its atomic weight. He found that it did not hold good when the temperature was near the melting point of the alloy.

245. *Specific Heat of Gases and Vapours.*—The specific heat of a gas, at any given temperature, may be measured in two different ways. We may keep the pressure constant, or we may keep the volume constant. The numerical values of the specific heat, as obtained by the two methods, are different; and so we speak of the *Specific Heat at Constant Pressure* and the *Specific Heat at Constant Volume.*

The experimental difficulties which are encountered in the determination of the latter are, by most methods (§ 243), almost insurmountable; but, in the case of approximately perfect gases, which closely obey the law $pv = Rt$ (§ 240), the principles of thermodynamics show (Chap. XXXI.) that the difference of the two specific heats is equal to the quantity R. Hence it is only necessary to measure the specific heat of such substances at constant pressure. The method which is adopted consists in passing a slow stream of the gas, under constant pressure, through two spiral tubes, in the first of which its temperature is raised to a known amount, while in the second it is lowered to a known amount. The amount of heat which is given out in the process of cooling is measured by the rise of temperature of a known mass of water, and the mass of the gas which gives out the heat is determined from a measurement of its volume.

Delaroche and Bérard, who first used the above method, were led to believe that the specific heat of a gas varies with its pressure. Regnault, who worked with improved apparatus, found that it is independent of the pressure—not merely in the case of a gas such as air which sensibly obeys Boyle's law, but also in the cases of carbonic acid and hydrogen. He found also that while the specific heat of carbonic acid increases markedly as the temperature rises, the specific heat of air is independent of the temperature. It is, therefore, by Boyle's and Charles' laws, independent of the volume. And we may in all probability conclude that the specific heats of all gases which closely obey Boyle's law are absolutely constant. It follows that the thermal capacity of such a gas is proportional to its density. The results in the table are due to Regnault.

Specific Heat of Simple Gases.

Hydrogen	3·4090	Oxygen	0·2175
Nitrogen	0·2438	Chlorine	0·1210
Air	0·2374	Bromine	0·0555

Specific Heat of Compound Gases.

Ammonia	0·5084	Carbonic acid ...	0·2169
Carbonic oxide ...	0·2450	Hydrochloric acid	0·1852
Sulphuretted hydrogen	0·2432	Sulphurous acid ...	0·1544

The principles of the kinetic theory of gases, on the assumption of hard smooth spherical molecules, acting by direct impact only, indicate that the specific heat should not vary with temperature. For the change of temperature is, on that theory, proportional to the change of mean square speed; and the change of kinetic energy is proportional to it also. Hence, the rate at which the energy varies with the temperature is constant. But when the action of molecular forces is considered, the rate may be either positive or negative.

By means of his steam calorimeter, Joly has verified Regnault's result that the specific heat of carbonic acid increases with the density. He finds (contrary to Regnault) an increase also in the case of air, and an apparent decrease in the case of hydrogen.

The *ratio* of the two specific heats of a gas may be found from the speed of sound in that gas (see § 134). In air and some other gases the specific heat at constant pressure is almost exactly 1·4 times greater than that at constant volume. Jamin and Richard have obtained, by a direct experimental method (in which the temperature of the gas is raised by means of a known amount of heat developed by the passage of an electric current through a metallic wire), results which agree well with those obtained by the acoustic method.

The specific heat of water vapour, under constant pressure, is about 0·48.

The specific heat of some saturated vapours—for example, those of water and carbon bisulphide—is *negative*. Such vapours, no liquid being present, become superheated under increase of pressure unless heat be withdrawn from them; and, under decrease of pressure, they will condense unless heat be supplied to them. Their specific heat diminishes in numerical magnitude as the temperature is raised. On the other hand, the specific heat of the saturated vapour of ether is positive, and increases with increase of temperature. In all cases the actual increase of specific heat is positive. In benzine this increase results in a change of sign of the specific heat at a definite temperature.

246. *Change of Molecular State. Latent Heat.*—We have already remarked that, in some cases, the application of heat to a body does not produce a rise of temperature, and that a change of molecular condition appears instead.

Thus the application of heat to ice below 0° C., raises its temperature and causes it to expand. When the ice reaches the temperature of 0° C., melting takes place continuously as more and more heat is applied. When the melting is complete the temperature again rises until, under ordinary atmospheric conditions, the water boils at 100° C. When all the liquid has boiled away the temperature again rises until the water begins to break up into its constituent elements.

If, at any stage of the above process, the application of heat were stopped, and heat were withdrawn instead, the various changes would be gone through in precisely the reverse order.

The change from the solid state to the liquid state is termed the process of melting or of *fusion*. The reverse process is called *solidification* or *regelation*.

The change from the liquid condition to the state of vapour is known as *vaporisation*, and the direct change from the solid state to the state of vapour is called *sublimation*.

In no case probably do these changes take place suddenly. Evaporation may go on at all temperatures; and many solids gradually soften before they melt. It is most probable that such softening occurs even in the case of ice and similar bodies which appear to melt suddenly.

The heat which is applied in order to produce fusion or vaporisation, without change of temperature, was called *Latent Heat* because it does not give rise to effects which can be measured by any ordinary thermometric apparatus.

247. *Fusion and Solidification.*—The laws which regulate the process of fusion, and which have already been alluded to in the case of water, may be enunciated as follows:

1. *So long as the pressure is maintained constant there is a definite melting-point for every solid;*

2. *If the solid and the liquid be well mixed, and heat be applied slowly, the temperature of the mixture remains at the melting-point until the whole of the solid has melted.*

In the statement of the first law the condition of constant pressure is imposed. Only in the case of a substance, the volumes of equal masses of which, in the solid and liquid conditions respectively, were equal, however the pressure might vary, would the restriction be unnecessary.

If a given liquid, such as water, expands in the act of solidification, the application of pressure will tend to prevent the solidification, *because* it tends to prevent expansion. Consequently, more heat must be withdrawn from the liquid in order that the change of state may be brought about. But this implies that the temperature is lowered.

Similarly, the melting-point (or, rather, from our present point of view, the solidifying-point) of a substance, such as paraffin, which contracts in the act of solidification, is raised by the application of pressure. For the application of pressure makes the change occur more readily; and, consequently, less heat has to be abstracted in order that the action may proceed. In other words, the temperature at which the change occurs is raised.

A theoretical investigation of the problem will be given later (Chap. XXXI.).

Professor James Thomson predicted from theory that the melting-point of ice would be lowered by pressure to the extent of 0°·0075 C. per atmosphere of pressure. This prediction was fully verified by Lord Kelvin. The second column in the table below gives the melting-points of paraffin, in Centigrade degrees, which correspond to the pressures, in atmospheres, which are given in the first column. These results were obtained by Bunsen. The second and third pairs of columns give similar results, obtained by Hopkins, for stearine and for sulphur respectively.

1	... 46°·3	1	... 72°·5	1	... 107°
85	... 48°·9	519	... 73°·6	519	... 185°
100	... 49°·9	792	... 79°·2	792	... 140°

The motion of glaciers is due, in large part at least, to the fact that the melting-point of ice is lowered by pressure. When the pressure arising from the weight of the superincumbent strata of ice increases to a sufficient extent at any point in the bed of the glacier, liquefaction takes place, and the water flows round the obstacle to the presence of which the increase of pressure was due. But, the pressure being relieved at the given point (and, therefore, handed on to another part of the mass) because of the contraction which takes place in melting, the water again becomes solid; and so the glacier, by a continuous process of melting and re-solidification, gradually moves down the valley which it occupies.

For the same reason, snow which is not too cold may be readily kneaded into a compact mass of ice; and a wire, which is loaded at its two extremities, and is hung over a bar of ice, will gradually cut its way through the bar without actually dividing it into two parts; for, though the ice below the wire is melted by the pressure, the water which is produced flows round the wire and solidifies above it. The path of the wire through the clear ice can be readily traced by means of the air-bubbles which the ice contains.

Lord Kelvin has found that the earth as a whole is more rigid than an equal globe of glass. This could be explained if the melting-point of the average materials of the earth is raised by pressure. It is well known that this is so in the case of ordinary lavas.

Under special circumstances, the laws of fusion, as enunciated above, may be violated.

Thus Fahrenheit showed that water which completely fills a closed glass vessel may be cooled below 0° C. before it freezes. And Gay Lussac showed that the temperature may be lowered to – 12° C. in an open glass vessel if the surface of the water be protected from the air by a layer of oil. The same phenomenon appears in the cases of other liquids, such as melted tin, phosphorus, and sulphur. In all such cases any vibration of the liquid must be avoided, else solidification will take place suddenly.

The melting-points of different substances vary greatly. On the one hand, hydrogen can only be solidified by the aid of powerful freezing mixtures; and, on the other hand, gas-coke can only be softened at the temperature of the electric arc.

Table of Melting-points.

Mercury ...	... −40°	Sulphur ...	... 115°
Ice ...	... 0°	Zinc	... 415°
Phosphorus	... 44°	Wrought iron	... 1500°(?)

The melting-point depends upon the purity of the substance. Thus the melting-points of different alloys of the same two substances vary greatly.

Rudberg showed that in an alloy the process of melting is gradual. The less fusible component melts first, and the amount of solidity of the whole then depends on the proportion of the fused component which is present. Complete fusion occurs only when the melting-point of the second component is reached.

In the case of alloys of lead and tin, the melting-points were found by Rudberg to be

$Pb_3 Sn$	... 289° C.	$Pb Sn_3$	... 186° C.
$Pb Sn$	... 241	$Pb Sn_4$	... 189
$Pb Sn_2$	... 196	$Pb Sn_5$	... 194

In the case of the compound $Pb Sn_3$, the lowest limit of temperature is reached, and so this substance is regarded as a chemical compound. As more lead was added, the melting-point rose till it reached the value found for pure lead; as more tin was added, it rose till the melting-point of tin was reached.

In general the temperature of fusion of the alloy is lower than that of either of the constituents. This result holds also in the case of fused salts.

248. *Latent Heat of Fusion.*—The latent heat of fusion of any substance may be defined as *the quantity of heat which is required to just melt one gramme of that substance at its ordinary temperature of fusion.*

This latent heat is given out again on re-solidification. Its amount, for each definite substance, under given conditions of pressure, is invariable.

The methods used for the determination of latent heat are essentially similar to those used for the determination of specific heat.

De la Provostaye and Desains, and Regnault, found the latent heat of fusion of ice to be equal to 79·25 units. Person, more recently, has used a different method, ultimately with the same result. He heated a quantity of ice, the temperature of which was originally below 0° C.; and, in consequence, he had to take account of the specific heat of ice, which, as we have seen, is about 0·5°. He at first obtained the value 80 for the amount of latent heat of fusion; but, subsequently, he traced the discrepancy between his result and that of previous observers to the fact that latent heat seems to be absorbed to a slight extent before the temperature 0° is reached. This, if true, furnishes evidence of the truth of the supposition, made in § 61, that the process of liquefaction is gradual.

The following values of the latent heat of fusion of some substances are taken from Person's results:

Latent Heat of Fusion.

Ice	79·25	Tin	14·25
Phosphate of soda ...	66·80	Lead	5·37
Zinc	28·13	Mercury ...	2·83

249. *Evaporation and Condensation.*—The laws of evaporation are similar to those of fusion.

1. *So long as the pressure is maintained constant, there is a definite boiling-point for every liquid.*

2. *If the liquid be well stirred, the temperature of both liquid and vapour remains at the boiling-point until all the liquid has evaporated.*

The effect of pressure is always in one direction with regard to the boiling-point, for all substances expand when they evaporate. The effect is therefore to raise the boiling-point, and its elevation is much more marked than is the alteration of the melting-point of a substance. But, before discussing this point farther, we must consider more fully the process which is termed *boiling*.

Evaporation occurs to a greater or less extent at all temperatures, and the rate of evaporation, *ceteris paribus*, increases rapidly as the temperature rises. If the liquid be contained in a closed vessel, the rate of evaporation gradually decreases and finally vanishes. It is not really true that evaporation has ceased. A state of kinetic equilibrium, in which the rate of evaporation is equal to the rate of condensation, has been attained. When this condition of equilibrium holds, the vapour is said to be *saturated;* and it is found that the pressure of the saturated vapour depends only on the temperature.

When the evaporation takes place in space of unlimited extent, the rate of evaporation from similar surfaces is proportional to the linear dimensions of the surfaces, and not to the total area. The theory of the subject has been worked out by Stefan. He finds that, in the case of a circular surface, the rate is proportional to the diameter, and the lines of flow of the vapour are hyperbolas, whose foci are on the circular edge. Winkelmann has verified this result in so far as his method of observation allowed. In the case of evaporation in a narrow tube, the rate is inversely proportional to the distance of the surface of the liquid from the end of the tube.

The presence of gases, such as air, has no influence upon the final state of equilibrium: it merely increases the time necessary for the attainment of the condition.

But the vapour may be saturated at any temperature as well as at the usual boiling-point. And this leads to the definition of the boiling-point as *the temperature at which the pressure of the saturated vapour is equal to that to which the free surface of the liquid is subjected.*

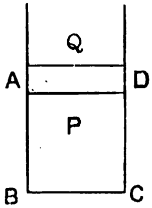


FIG. 136.

The following remarks should make the matter clear. Let ABCD (Fig. 136) represent a cylinder in which a smooth, massless (and therefore weightless) piston AD, which we also suppose to be gas-tight, works freely. First, let there be a gas in the closed region P and another gas in the region Q outside the piston. Evidently equilibrium is only reached when the pressure is the same on both sides of the piston. Now suppose that P is filled with a liquid below its boiling-point. The filling of the region P is necessarily complete so long as vapour is not formed. And no vapour can be formed until the pressure of that vapour is equal to the pressure of the gas in the region Q, *i.e.*, until its pressure is equal to the pressure to which the free surface of the liquid is exposed. But when the liquid is at the temperature at which this occurs, vapour will be formed, and the continued application of heat will force the piston up. If we now suddenly produce a vacuum in the region Q, vapour will be rapidly formed in P; and that vapour will proceed not merely from the surface of the liquid, but also in bubbles from its interior. This process of free evaporation is called *boiling* or *ebullition.*

The phenomena exhibited by Geysers are due to a like cause. A sudden reduction of pressure, or an increase of temperature, in the interior of the column of water which fills the funnel causes the water at that part to change its state explosively, and so the superincumbent water is ejected violently.

The following table, given by Regnault, exhibits the relation between the boiling-point and the pressure :

Pressure of the Saturated Vapour of Water.

Temp. C.	Pressure in Atmospheres.	Temp. C.	Pressure in Atmospheres.
0°	0·006	120°	1·962
10°	0·012	130°	2·671
20°	0·023	140°	3·576
30°	0·042	150°	4·712
40°	0·072	160°	6·120
50°	0·121	170°	7·844
60°	0·196	180°	9·929
70°	0·306	190°	12·425
80°	0·466	200°	15·380
90°	0·691	210°	18·848
100°	1·000	220°	22·882
110°	1·415	230°	27·535

By sufficiently reducing the pressure, water may be made to boil violently—not merely to evaporate at its surface—at temperatures far below its ordinary boiling-point. The well-known experiment of causing hot water to boil in a closed flask, by pouring cold water upon the flask, is a case in point. The sudden reduction of temperature causes partial condensation of the vapour already formed in the flask, and so gives rise to a sudden diminution of pressure.

When the boiling-point is known, the atmospheric pressure may be obtained from a table such as that above. The *Hypsometric Thermometer*, used for the determination of height above sea-level, is based upon this principle. The atmospheric pressure diminishes as the elevation above sea-level increases, and the result is that the boiling-point is lowered by about 1° C. at an elevation of 960 feet above sea-level.

In the case of a real or apparent mixture of liquids, Regnault found that the vapour pressure of the mixture is equal to the sum of the vapour pressures of the components if, and only if, the liquids do not really mix. If the liquids mix, the vapour pressure is less than the sum of the vapour pressures of the components, and sometimes is less than that of one component. The more readily the constituents mix, or the greater is their 'affinity,' the greater is the defect of the vapour pressure from the sum.

The law that the total pressure is the sum of the partial pressures is known as *Dalton's Law*, and is true of a mixture of gases which have no chemical action upon each other. This results from Boyle's law. For the sum of the products of pressure and volume is necessarily equal to the product of the total pressure into the volume, and the same volume is filled by each gas. Hence the sum of the pressures is equal to the total pressure. But we cannot expect that the law will hold for vapours which do not, to a sufficient approximation, obey Boyle's law. Even if Dalton's law is followed, it does not necessarily result that the vapour densities of the components have their normal values. That this is so was proved experimentally by Regnault.

The laws of evaporation are subject to exceptions, just as are the laws of melting. Thus water may, by cautious heating in a smooth clean glass vessel, be raised considerably above its ordinary boiling-point, if it has been carefully freed from dissolved gases. A very slight vibration may then cause it to boil explosively. Tomlinson's investigations on this subject show that air introduced into liquids at their boiling-points does not promote boiling unless it be dusty, and indicate that a surface which is chemically clean is inactive. Non-porous substances become inactive after they have been used to promote boiling for some time. But this does not result with porous substances, such as carbon, etc. Porous substances should, therefore, be used to prevent explosive boiling.

The boiling-points of various liquids differ greatly under ordinary atmospheric conditions, as the following table shows:

Table of Boiling-points of Liquids.

Zinc	1040° C.	Bisulphide of carbon ...	48° C.
Mercury	350°	Sulphurous acid	−10°
Water...	100°	Nitric oxide	−87°

250. *Latent Heat of Vaporisation.*—Regnault found for the 'total heat of steam,' *i.e.*, the quantity of heat which is given out by one gramme of water in condensing to water at 0° C., the expression,

$$H = 606{\cdot}5 + 0{\cdot}305t,$$

where t is the temperature in Centigrade degrees. The latent heat is obtained by subtracting from the total heat the quantity of heat which is given out by the liquid in cooling from $t°$ C. to 0° C. Hence we get for the latent heat the expression

$$L = 606{\cdot}5 - 0{\cdot}695t - 0{\cdot}00002t^2 - 0{\cdot}0000003t^3.$$

This formula is true throughout the range of temperature from 0° C. to 230° C. If we could assume that it held true up to 706° C., it would indicate that the latent heat vanishes at that temperature very nearly. (See § 252.)

Trouton has shown that, to a very close degree of approximation, the molecular latent heat is proportional to the value of the normal boiling-point on the absolute scale.

The latent heat of steam is very large in comparison with that of most other liquids, as this table shows:

Latent Heat of Vaporisation.

Water	536	Ether	90·4
Naphtha	264	Bisulphide of carbon ...	86·7
Alcohol	202	Bromine	45·6

We have found previously that the latent heat of liquefaction of ice is also relatively large. These facts—of the large latent heats of liquefaction of ice and of vaporisation of water—are of great importance in the economy of nature. If they were not so, destructive floods might frequently occur from the rapid liquefaction of ice, or sudden condensation of moisture, consequent on a slight variation of temperature.

The latent heat of vaporisation is used for the production or maintenance of low temperatures. Water may be kept cool in very hot weather if it is enclosed in a vessel of porous earthenware; for part of it percolates through the vessel and evaporates from its outer surface, the latent heat being largely drawn from the vessel and its liquid contents. Also solid carbonic acid is produced if a jet of the liquid (formed under considerable pressure) is allowed to escape from the vessel which contains it; for the outer parts of the jet evaporate, and the necessary latent heat is largely taken from the interior parts of the jet, which consequently are solidified. Faraday froze mercury in the interior of a white-hot platinum crucible, by placing it in a capsule which rested on a mixture of solid carbonic acid and ether contained in the crucible.

251. *Formation of Dew.*—When a superheated vapour is cooled sufficiently, saturation takes place, and any further cooling causes condensation. The moisture which is deposited in this way from the atmosphere is termed *dew.* Any cold body lowers the temperature of the air in immediate contact with it; and, when the temperature is sufficiently lowered, a thin film of moisture is deposited upon the cold body. The latent heat which is given out on condensation gradually raises the temperature of the cold body until it becomes equal to that which corresponds to the vapour-pressure, at which stage the action ceases. This temperature, being also that at which the deposition of dew just commences, is called the *Dew-point. Hoar-frost* is formed when the dew-point is below 0° C.

Wells first gave the correct explanation of the formation of dew. He showed that dew is freely deposited on nights when the sky is clear, because on such nights the earth loses heat rapidly by radiation, and so cools rapidly to the dew-point; whereas, on cloudy nights, the clouds absorb, and radiate back to the earth, a large part of the radiated heat, so that the ground does not cool rapidly. Another condition necessary to the ready formation of dew is that the air shall be still, otherwise no portion of the air may remain in contact with the ground for a length of time sufficient to allow of its being cooled to the dew-point. The dew will, of course, deposit itself most freely on those bodies which part with their heat most rapidly and have also small specific heat.

Aitken has recently shown that the presence of particles of dust is in general necessary before condensation of moisture can occur in the atmosphere, and that supersaturation of a vapour can be produced by getting rid of all dust-particles by filtration through cotton-wool. These particles act as nuclei upon which the deposition takes place. This phenomenon is very closely connected with the phenomenon of the dependence of the equilibrium-pressure of vapour upon the curvature of the liquid film with which it is in contact (§ 103).

Aitken has also utilised the fact that moisture is deposited upon the dust-particles in the construction of an instrument which enables us to determine the number of particles which are contained in a given volume of any definite specimen of air. This instrument is certain to prove of considerable meteorological importance.

Daniell's *Hygrometer* was constructed for the purpose of accurately registering the dew-point. It consists of two hollow glass bulbs connected by a glass tube. One of these bulbs is made of black glass, and the other is made of clear glass. A small thermometer, the stem of which projects into the tube of (clear) glass which connects the two bulbs, is placed in the black bulb along with a quantity of sulphuric ether. The remaining portions of the interior of the instrument are filled only with the vapour of ether. A piece of cambric is tied round the other bulb, and a little ether is poured upon it. The evaporation of this ether cools the bulb, and makes some of the vapour inside it condense. This destroys the equilibrium of the liquid ether and its vapour in the interior of the instrument; and some of the ether in the black bulb evaporates, in order that equilibrium may be restored. The absorption of latent heat cools this bulb, and, finally, dew is deposited on its exterior. The presence of a very slight film of dew is readily observed on the black surface, and the temperature of the bulb is noted; but the reading of the thermometer is necessarily a little too low. The evaporation is then stopped, and the temperature at which the dew just disappears is observed. This reading is a little too high, and so the mean of the two results is taken.

Regnault introduced improvements which rendered it possible to observe the appearance and disappearance of the dew at practically one temperature.

His apparatus consists essentially of a wide glass tube, the lower end of which is closed by a thimble-shaped piece of thin polished silver, inside of which is placed some ether. The other end of the glass tube is closed by a cork, through which passes a thermometer whose bulb is immersed in the ether. A small open glass tube also passes through the cork and reaches down to near the foot of the silver end. A side tube leading from near the top of the wide tube is connected to an aspirator, by means of which air is drawn through the apparatus. The bubbling of the air through the ether causes the latter to evaporate, and so lowers the temperature, and dew is ultimately deposited on the outside of the silver. A second tube, similar to that containing the ether, but containing air alone, is connected to the apparatus, and a thermometer inside it registers the temperature of the air. The bubbling of the air through the ether keeps the whole mass of liquid uniformly at a temperature which is practically that of the thin metal end. The aspirator may be so regulated that the dew appears and disappears at practically the same temperature.

The dew-point is also found by means of *Wet and Dry Bulb Thermometers*. The one thermometer has its bulb surrounded by cambric, which is kept moist with water drawn up by capillary action through some threads which dip into a vessel containing it. The other (ordinary) thermometer registers the exact temperature of the air. The reading of the wet-bulb thermometer is lower than that of the dry-bulb thermometer so long as evaporation is going on; but, if the atmosphere is saturated with water-vapour, no evaporation takes place, and both thermometers register the same temperature. The formula

$$p = p_0 - \frac{\delta}{48} \cdot \frac{b}{30},$$

in which p represents the pressure of water-vapour in the atmosphere, p_0 represents the pressure which is given in Regnault's table as corresponding to the temperature of the wet-bulb, δ is the difference between the wet-bulb and the dry-bulb readings, and b is the height of the barometric column in inches, was found by Apjohn to accord well with observed results.

252. *Continuity of the Liquid and Gaseous States. Critical Temperature.*—Cagniard de la Tour first showed that a substance may exist in a non-liquid state at a density very nearly equal to its density in the liquid condition. A complete investigation of the subject was made by Andrews, who showed that

There is a Critical Temperature for every vaporous or gaseous substance, such that no amount of pressure can liquefy the substance, unless its temperature be below the critical value.

The critical temperature of carbonic acid is 30°·9 C. That of water is about 412° C.

The latent heat vanishes at the critical temperature. We have already seen that the latent heat of water should vanish at about 706° C., if Regnault's formula connecting latent heat with temperature held throughout that range. The result just given shows that the formula deviates largely from the truth at temperatures higher than the limit (230° C.) up to which Regnault worked.

The accompanying diagram (Fig. 137) represents the results of Andrews' experiments on carbonic acid. Pressure is measured (in atmospheres) along the axis of ordinates, and volume is measured along that of abscissæ. At the temperature 13°·1 C., the volume of the gas gradually diminishes, as the pressure is raised, until liquefaction commences. After this, the volume lessens, without any rise of pressure, until all the substance is liquefied; and then immense pressure is required to lessen it even slightly.

Similar effects take place at the higher temperature 21°·5. The line (called an *isothermal*) which represents the simultaneous values of pressure and volume at this higher temperature, lies, in the diagram, entirely to the right hand of, and above, the isothermal of

13°·1; for, the pressure being constant, the volume increases with the temperature, and, the volume being constant, the pressure increases with the temperature. But, at this higher temperature, the change of volume in passing from the gaseous to the liquid state is smaller than that which occurs at the lower temperature; the liquefaction commences at a smaller volume, and ends at a larger volume, than when the temperature is less. The isothermals cease to have a portion parallel to the axis of volume, *i.e.*, liquefaction ceases, at 30°·9.

The dotted curve separates the region in which the liquid and the vapour can exist together in equilibrium from the regions in which the substance is entirely liquid or entirely vapour. The isothermal of 30°·9 separates the region in which liquefaction can occur from that in which it is impossible.

We may with great advantage, as Tait suggests, describe the substance as a *true gas*, or a *true vapour*, according as the temperature is higher, or lower, than the critical temperature.

The compressibility of the substance is v/Vp, where V is the volume and v, p, represent respectively simultaneous small alterations of the volume and the pressure. Now, the diagram shows that, at the commencement of liquefaction, the inclination of the isothermal to the axis of volume becomes greater and greater as the temperature rises, *i.e.*, the ratio v/p decreases as the temperature rises. Hence, since we suppose unit volume to be taken in all cases, the compressibility of the vapour when it is upon the point of condensing decreases as the critical temperature is approached. Similarly, the value of v/p in the liquid state, when the substance has just been entirely liquefied, increases as the temperature rises; and thus we see that the compressibility in the two states tends towards equality, simultaneously with the volumes, as the temperature rises to its critical value.

For a considerable distance above the critical point the isothermals exhibit two points of inflexion; but these finally cease to be visible, and the isothermals closely resemble those of a perfect gas.

[The illustration affords a good example of the use of contours. The isothermals may, as was stated in Chap. III., be regarded as the projections of the plane sections of a surface which represents the various simultaneous value of the pressure, volume, and temperature of the gas.]

253. *Solution. Freezing Mixtures.*—The process of solution is extremely analogous to the processes of liquefaction. A gas which is dissolved in a liquid may be regarded to a certain extent as if it were liquefied, and latent heat is given out in the process of solution.

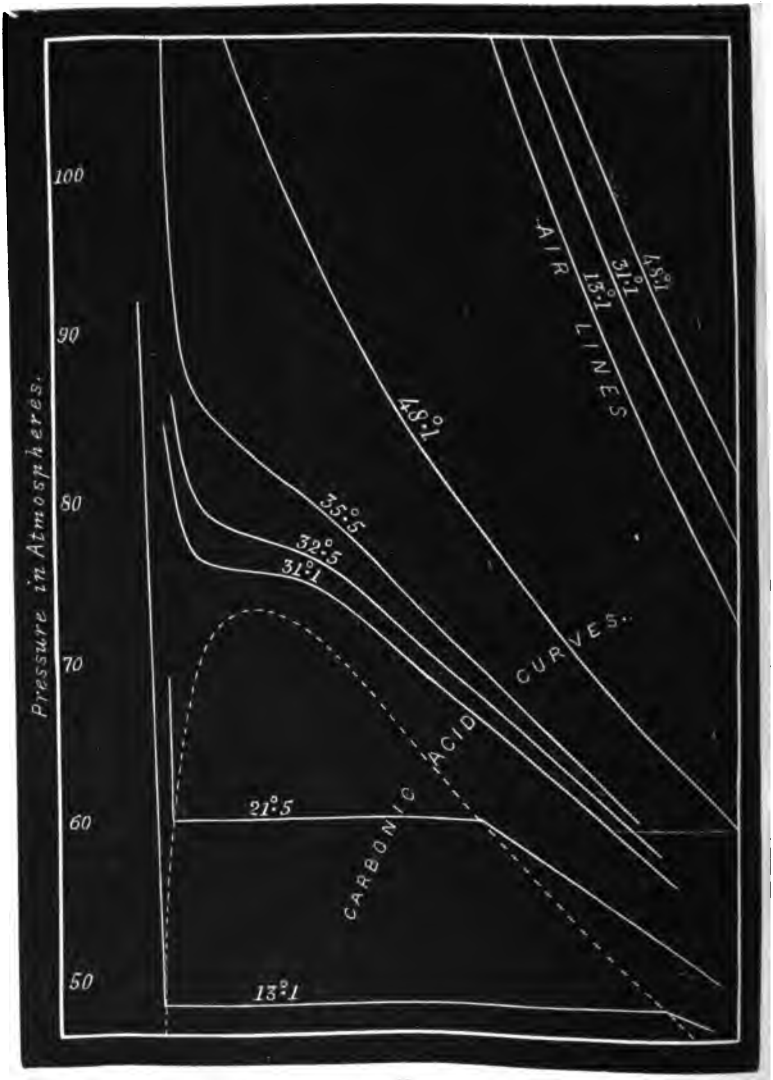

Pressure in Atmospheres.
100
90
80
70
60
50
48°1
31°1
13°1
AIR LINES
48°1
35°5
32°5
31°1
CARBONIC ACID CURVES.
21°5
13°1

FIG. 137.

Similarly, when a solid is dissolved in a liquid, latent heat is required. just as if the solid were directly liquefied; but in some cases this absorption of latent heat is masked by the heat which is developed because of molecular action between the liquid and the solid.

The amount of heat which is disengaged in the solution of a gas is frequently very marked. This is so specially in the cases of the more soluble gases, such as ammonia when water is the solvent.

The amount of gas, under definite pressure, which a given liquid will dissolve, becomes less as the temperature is raised; though, by careful treatment, a state of supersaturation may be induced—which is analogous to the prevention of boiling at temperatures considerably over the ordinary boiling-point of a liquid.

Supersaturation of a liquid solution of a solid may also take place —notably in the case of a substance, such as acetate of soda, which dissolves in little more than its own water of crystallisation. If a crystal of the acetate be dropped into the supersaturated solution to act as a nucleus, crystallisation will take place rapidly with the development of latent heat. A crystal of any other substance of the same crystalline form will produce the same effect.

Heat is frequently developed or absorbed when two liquids are mixed (mutually dissolved). If chemical action takes place to any extent between the two, heat will be developed unless other causes prevent. If the total bulk of the two liquids increases on mixture— as in the case of bisulphide of carbon and alcohol—heat tends to be absorbed; and again, the water-equivalent of the mixture may be greater than the sum of the water-equivalents of its constituents— which also necessitates absorption of heat. If the opposite effects to these take place, heat will be evolved. Energy may also be changed into heat in the process of inter-diffusion of the liquids. Disengagement or absorption of heat takes place respectively according as the effects of the one or the other sets of actions preponderate. And, as the various actions depend upon the temperature, we find that the total effect is sometimes reversed when the original temperature of the two liquids is sufficiently varied.

Two solids even may dissolve in each other, so to speak, with the absorption of latent heat. (Salt and snow furnish a well-known example.) This can obviously only occur when the freezing-point of the resultant liquid is lower than the original (common) temperature of the solids. Part of the latent heat is obtained by cooling the solids, part by cooling the liquid, and part, it may be, by cooling surrounding bodies. If the whole be intimately mixed, the action necessarily ceases when the freezing-point of the resultant liquid is reached. These remarks contain the explanation of the action of solid freezing-mixtures, which has been elaborately investigated by Professor Frederick Guthrie.

254. *Dissociation and Chemical Combination.*—When the temperature is raised sufficiently high a compound dissociates, or breaks up, into its constituents. The change is not sudden but gradual. It commences at a certain lower limit of temperature, and ends completely at a certain higher limit; and at all intermediate temperatures a state of kinetic equilibrium is arrived at in which recombination precisely balances dissociation. The magnitude of the limits will, in general, depend upon the pressure. It is usual to speak of the temperature at which one-half of the substance is dissociated as the temperature of dissociation.

Conversely, when the two (or more) constituents are mixed, combination does not occur until a certain temperature is attained; but, if the combination results in the development of heat, the process, once started, will continue until the percentage of the mixture which remains uncombined corresponds to the temperature which the whole mass attains because of the heat which is set free. On the other hand, if work is done during the process, or if heat is lost by conduction or otherwise, the process will continue until combination is complete, when the temperature falls to the lower limit.

All chemical combination takes place in accordance with the two laws of thermodynamics, and therefore further treatment of this subject is deferred until we have considered these laws. (See § 272.)

255. Many other effects of heat might be noted here, but it is preferable to leave their discussion to those special sections in which we have to treat of the properties which are affected by the application of heat.

CHAPTER XXII.

CONDUCTION AND CONVECTION OF HEAT.

256. *Conduction.*—We have already discussed the transference of heat by the process of radiation, that is, the transference of heat without the mediation of ordinary matter. In the process of radiation the transferred energy may pass *through* a material substance without being communicated to it; indeed, radiation ceases in so far as such communication is made. We must now consider its transference when ordinary matter is the medium through which it is transferred.

The most marked difference between the two cases lies in the rate of propagation, which is extremely rapid when radiation occurs, while it is extremely slow in comparison when ordinary matter is the medium of transference.

Two methods exist according to which heat (which consists in kinetic energy of molecular motion) may pass from one place to another by means of matter. The energy may pass from one portion of matter to another, which occupies a different position, by actual (or virtual) impact between the molecules of the two portions; that is to say, it may be handed on from one portion to another: or again, it may pass, not from one portion of matter to another, but from one locality to another by motion of the hot body. The former of these processes is known as *Conduction;* the latter as *Convection.* Both take place in liquids and in gases; the former alone can take place in solids. Of course, internal radiation may have, and very likely has, a considerable, and possibly an essential, part to play in the process. Yet the immense difference between the rates of transference of heat by radiation in free space and by conduction through a solid makes the intervention of ordinary matter *the* feature of the process.

257. *Conductivity.*—Different substances, under like conditions, conduct heat at different rates. Thus a bar of iron, one end of

which is red-hot, may be too hot to grasp at the cooler end; while a bar of wood, of the same length, which is burning at one end, may be easily handled at the other. The property in virtue of which such differences arise is termed *Conductivity.*

Most experiments which are intended to illustrate the differences between the conductivities, or conducting-powers, of various substances for heat exhibit only the differences between the rates at which the temperatures of the substances, at a given distance from the source of heat, attain a definite value under given conditions. The well-known experiment of Ingenhausz is of this description. In it, a series of similar, uniform, and equal, rods, of different substances, project from the side of a metallic trough into which hot water is suddenly poured. Each rod is coated with a thin film of beeswax which melts at a definite temperature. The rate at which this definite temperature travels along each rod is plainly shown by the motion of the line of demarcation between the melted and the unmelted portions of the wax. But obviously this rate will only coincide with the rate at which heat is conducted along when the thermal capacities of the various substances are practically identical, for, other things being equal, the rate at which the temperature rises is inversely proportional to the thermal capacity.

Fourier was the first to give an accurate definition of conductivity. The whole subject of heat-conduction was so fully and accurately developed by him that his work—'Théorie analytique de la Chaleur' —published in 1822, still remains the text-book on the subject.

Let us suppose that the substance, the thermal conductivity of which we are considering, is in the form of a uniformly thick plane slab of practically infinite extent. Let θ be its thickness; and let one side be kept at a uniform temperature t, while the other is kept at a uniform temperature t' until a steady flow of heat takes place from side to side. The quantity of heat, h, which passes in τ units of time through an area a of the surface of the slab is found experimentally to be directly proportional to τ, a, and $t'-t$, while it is inversely proportional to θ. Hence we may write

$$h=k\,a\frac{t'-t}{\theta}\tau$$

The quantity $(t'-t)/\theta$ is called the *temperature gradient,* and k is the conductivity.

If the area, the temperature gradient, and the time, be all unity, the equation becomes

$$h=k,$$

and so we obtain the following definition of the conductivity:

The thermal conductivity of a substance, at any temperature, is the number of units of heat which pass, per unit of time, through unit of surface of an infinite slab of the substance, of unit thickness, the sides of which are kept respectively at temperatures half a degree higher, and half a degree lower, than the given temperature.

In making this definition we assume that the unit of length is not excessively small—that it is (say) a centimetre, an inch, or a foot—and that an ordinary temperature degree—say, the Centigrade—is used, so that the temperature gradient is not large. The necessity for these restrictions is apparent, if we consider that the conductivity may (it actually does) vary somewhat with the temperature; for, in consequence of such variation, the temperature gradient could not be sensibly uniform from side to side of the slab, if the difference of the temperatures at the two sides were large. As a special case, let us suppose that the conductivity of a layer of the slab of half its total thickness, is one-half of that of the remaining portion. Since the same flow of heat takes place through both portions, the difference of temperature between the sides of the former portion must be double of that between the sides of the latter.

Of course, even if the conductivity varies from point to point, whether from variation of temperature or from any other cause, the quantity k, determined from the above formula, will always represent the average conductivity of the slab considered as a whole.

But, quite apart from the question of such variation, we cannot assert that the quantity of heat which will pass through a slab, one unit in thickness, under unit difference of temperature, will be precisely equal to the quantity which will pass through a slab, n times thinner, under a difference of temperature n times less, when n is a very large number, and all the other conditions are unaltered.

258. *Measurement of Conductivity.*—Estimates of the relative conductivities of substances may be made by means of Ingenhausz's apparatus if the quantity measured in each case be the maximum distance along the bar to which the melting of the wax can take place. The conductivity is proportional to the square of this distance. The question of difference of thermal capacities does not arise in this case, since the temperature has attained a steady state, but difference of surface emissivities affects the result.

The relative conductivities may also be determined by means of the readings of three equidistant thermometers. This method was employed by Depretz, who coated the surfaces with lampblack in order to make the emissivities the same. But this had the result of making the emissivities so great that the temperature practically ceased to vary at no great distance from the heated ends of the bars. The bulbs of the thermometers were inserted in three equidistant holes drilled in the bar, and good contact was ensured by surrounding them with mercury or a fusible alloy. The results are affected by any want of homogeneity in the material of the bars. Wiedemann and Frantz improved the method by using electro-plated bars. This secured uniformity and smallness of surface emissivity. They also used a movable thermopile to register the temperatures.

In the earlier attempts to obtain absolute measurements of conductivities, one face of a plate of the substance was heated by steam, and the other was kept cool by means of ice or water. The great defect of the method lies in the assumption that the faces of the plate attain the temperature of the substances which are placed in contact with them—an assumption which is far from being true. This method has recently been greatly improved—the temperature being found at points inside the material of the plate, and the steam and ice being applied over an area considerably greater than that across which the flow of heat is estimated. The quantity of heat is found by weighing the amount of steam which is condensed, or of ice which is melted, by contact with the limited area. Berget introduced the farther improvement of using the vessel containing the ice, which is applied to the limited area, as a Bunsen's ice-calorimeter, in which the quantity melted is determined by its change of volume. He found the following values of conductivities, expressed in centimetre-gramme-second units:

Mercury ...	0·02015	Brass ...	0·26250
Iron ...	0·15870	Copper ...	1·04050

In another form of experiment for the absolute determination of conductivity, a steady, though not uniform, state of temperature is maintained throughout the substance. This method was used by Lambert, and subsequently, under greatly improved conditions, by Forbes.

In Forbes' method a long bar of the substance of uniform cross-sectional area is used. One extremity of the bar is inserted in a bath of melted lead, or solder; and the other extremity is exposed to the air, or, if necessary, is cooled by a current of water. Small holes, into which a little mercury is poured, are drilled in the bar at regular intervals; and these holes are lined with iron (if the bar itself be not made of iron) in order to prevent amalgamation. Thermometers, inserted in the holes (which are found not to appreciably affect the flow of heat along the bar) register the temperature of the bar in their immediate vicinity.

If distance measured along the bar from the source of heat be laid off along *ox* (Fig. 138), and if ordinates be drawn at points such as *p*, corresponding to the positions of the thermometers, and of lengths which are proportional to the readings of the thermometers at these points, a curve drawn free-hand through the extremities of the ordinates will enable us to obtain the temperature gradient at any part of the bar. For the tangent of the angle of inclination of the line which touches the curve at the extremity of any ordinate is equal to the space rate at which the temperature varies, per unit of length, at the corresponding section of the bar; *i.e.*, it is equal to the gradient of temperature at that section. But the sectional area of the bar is known, and hence, if we can determine the quantity of heat which passes in a given time through the given section, we can determine the conductivity by calculation from the equation above.

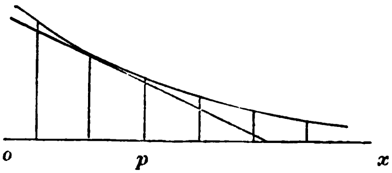


FIG. 138.

Now the heat which passes any section is entirely lost from the remaining portion of the bar by radiation, or otherwise; and any heat which is given to the water employed in cooling the far end of the bar, if this is required, can be readily estimated by means of the rise in temperature of the water, while the heat which is lost by radiation and convection is found by a special experiment.

During the above experiment, a thermometer, inserted in a hole in a small bar which is cut originally from the long bar, indicates the temperature of the air in the neighbourhood of the large bar. In the second experiment, which is made for the purpose of determining the rate of loss of heat, the small bar is heated uniformly to a temperature higher than the highest recorded in the former one. The bar is now allowed to cool, and the thermometer which is inserted in it enables us to determine the rate of loss of heat, per unit of time, per unit of length of the bar (§ 243). The mass of unit length is known, the specific heat is also determined, and the product of these quantities into the rate of fall of temperature gives the rate of loss of heat. This being known for all the various temperatures observed at the different parts of the bar in the first experiment, the total rate of loss of heat from the portion of the large bar, beyond any given section, is easily calculated. And so the thermal conductivity, at particular temperatures, can be found.

In the second experiment the temperature of the air is obtained by means of the long bar, so that the results in both cases can be compared, as is necessary, at the same excess of temperature over that of the surrounding air. But in addition to this, since the rate of cooling depends upon the temperature and pressure of the air, it is necessary to perform both experiments under as nearly as possible the same conditions of temperature and pressure.

The unit of heat which is employed in this method is ordinarily the amount of heat which is required to raise the temperature of unit volume of the substance by one degree, the amount of the heat being measured in terms of changes of temperature in the bar. Consequently the quantity which is so determined is not the thermal conductivity as above considered. Maxwell calls it the *Thermometric Conductivity;* Lord Kelvin calls it the *Thermal Diffusivity.* The thermal conductivity of any substance is obviously the product of the thermometric conductivity into the thermal capacity of that substance.

Tait, who repeated and extended Forbes' experiments, gives the following values of

Thermometric Conductivity.

Temperature C.	0°.	100°.	200°.	300°.
Iron	0·0149	0·0128	0·0114	0·0105
Copper, electrically good	0·076	0·079	0·082	0·085
Copper, electrically bad ...	0·054	0·057	0·060	0·063
German silver	0·0088	0·009	0·0092	0·0094

This table indicates that, with the exception of that of iron, the thermometric conductivity of all these substances increases as the temperature rises.

Tait also gives, for the iron and the two specimens of copper (the units being the foot, the minute, and the degree C.), the following values of

Thermal Conductivity.

Iron	0·788 $(1-0·00002t)$
Copper, electrically good ...	4·03 $(1+0·0013t)$
Copper, electrically bad ...	2·84 $(1+0·0014t)$

It appears, therefore, that in general the thermal conductivity increases as the temperature rises. Also, the order of the metals with regard to conduction of heat is the same as their order with regard to conduction of electricity. Forbes had observed this fact, and had expected that, as in the case of electric conduction, the thermal conductivity would decrease as the temperature becomes higher. This, as we see, does not, in general at least, hold true.

Dr. A. C. Mitchell has recently, under Professor Tait's direction, repeated these experiments with the same bars, but under improved conditions. For one thing, all the bars were nickel-plated, so as to avoid alterations of the surface from oxidation at high temperature. His results are, on the whole, confirmatory of the previous results —with this chief exception, that he found that the temperature coefficient for iron may be positive, as it is in all the other substances.

A different method was employed by Angström. In his method one end of the bar is alternately heated and cooled during equal periods of time, the temperature of the source of heat being kept constant. The alternations of heating and cooling are maintained until all the thermometers indicate practically periodical changes of temperature. Fourier's mathematical investigations show that, if the variations of temperature do not sensibly affect the conductivity and the specific heat, the conductivity can be calculated from the rate at which the range of temperature diminishes per unit of length of the bar, together with the observed speed at which the 'waves of temperature' run along the bar, provided that the rate of surface-loss of heat is proportional to the excess of the temperature of the bar over that of its surroundings.

259. *Conduction through the Earth's Crust.* — Angström's method has a direct application to the problem of the conduction of the diurnal and annual waves of solar heat downwards through the crust of the earth. In this investigation we may assume that the heated surface is practically an infinite plane, and that the flow of heat takes place in lines perpendicular to this plane.

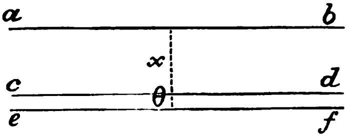


FIG. 139.

Let *ab* (Fig. 139) represent the surface, and let *cd*, *ef*, represent planes parallel to the surface, at distances x and $x+\theta$ respectively, from it. If c be the thermal capacity of the substance through which the flow takes place, while r is its rate of increase of temperature, the quantity of heat which enters a portion of the substance, of small thickness θ and area a, in a small interval of time τ is

$$car\tau\theta, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

for $ca\theta$ represents the quantity of heat which must be added in order to increase the temperature of the volume $a\theta$ of the substance by one degree, and $r\tau$ is the change of temperature in the time τ.

If k be the conductivity of the substance, supposed to be constant, while g is the temperature gradient measured downwards, the quantity of heat which, in the time τ, crosses in the positive direction the area a of the side of the slab nearest the surface is

$$-a\tau kg.$$

Similarly, the quantity which passes downwards through the area a of the surface, which is distant from the former by the amount θ, is (since θ is small)

$$-a\tau k(g+g'\theta),$$

where g' is the rate at which the gradient varies with depth. Consequently, the amount of heat which, on the whole, enters the volume $a\theta$ in the time τ is

$$a\theta\tau kg', \quad . \quad . \quad (2)$$

and hence we get

$$cr=kg', \quad . \quad . \quad (3)$$

since each of the quantities (1) and (2) represents the same amount of heat.

Now (§§ 22, 49) let us consider this equation simply as an equation of dimensions. The quantity r involves temperature directly and time inversely. The quantity g' involves g directly and distance inversely, *i.e.*, it involves temperature directly and the square of distance inversely. We get, therefore,

$$\frac{c}{t}=\frac{k}{l^2}, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

We must remember that the sign of equality indicates now equality of *dimensions* alone.

From (4) we get

$$l=\sqrt{\frac{kt}{c}},$$

which means that, if the times are altered in any fixed proportion p, the lengths must be altered in proportion to the square root of p

in order that the flow of heat may take place under *similar* conditions in the altered circumstances. In other words, *the distances at which similar effects are felt (for example, the distances below the surface at which periodic variations of surface-temperature, of given magnitude, cease to be felt) are proportional to the square root of the period.*

Now the period of the annual heating and cooling is 365 times as great as the period of diurnal variations. Hence, neglecting difference of range at the surface, the effect of the summer's heat is felt about nineteen times as far below the surface as the effect of the diurnal heat is felt.

Again,

$$\frac{l}{t}=\sqrt{\frac{k}{ct}}$$

Here we may suppose that l represents the length of a wave, while t is the periodic time, so that the fraction on the right hand is proportional to the rate at which the wave of heat travels downwards. We see, therefore, that this rate is directly proportional to the square root of the conductivity, and is inversely proportional to the square root of the thermal capacity and the periodic time conjointly.

It follows that, when the period is constant, *the date at which the maximum temperature reaches any given depth is later than the date at which it left the surface in direct proportion to the depth.*

The law which regulates the diminution of the range of temperature with increase of depth cannot be obtained from equation (3) in the way in which we have obtained the two laws just enunciated; for the temperature does not appear in equation (4). But if the heat-wave travelled downwards without change we could write (3) in the form

$$cgv=kg',$$

where v is the speed with which the heat-wave travels downwards. For the substitution of gv for r is equivalent to an assertion that the change of temperature at any point in a given, very small, time is due to the downward motion of the wave of temperature. Thus (3) would become

$$\sqrt{\frac{ck}{\mathrm{T}}}\cdot g=kg',$$

where T is the periodic time. Now this equation is not applicable to the actual heat-wave, since it does not travel downwards without change. But we may fix our attention upon, say, a point of maximum temperature, and follow it in its downward motion; and we may regard this varying maximum as a wave which moves relatively to that point regarded as fixed. In this case the previous reasoning applies rigorously, and the equation is strictly correct, providing only that we regard g and g' as referring to the change of the maximum temperature.

The equation asserts that the rate of diminution of the rate of change of temperature with depth is proportional to the rate of change itself. In other words, the rate of change diminishes in geometrical progression as the depth increases in arithmetical progression. And its rate of diminution is $\sqrt{c}/\sqrt{kT}$. But, since the rate of diminution of the rate of alteration of the range is proportional to the rate of alteration itself, it follows that the rate of alteration bears the same ratio to the range. Hence *the range diminishes in geometrical progression as the depth increases in arithmetical progression*, the rate of diminution being directly as the square root of the thermal capacity, and inversely as the square roots of the conductivity and the periodic time conjointly. It must, of course, be remembered that this investigation refers only to a wave of given period, for which accordingly the downward speed is constant.

As the result of direct observations (begun by Forbes in Edinburgh in 1837) of the temperature at different distances below the surface of the earth, it is found that the annual heat-wave travels inwards at the rate of little more than sixty feet per annum, and that the range of temperature has diminished to a very small fraction of its original amount when half of that distance has been traversed. The diurnal heat is therefore inappreciable at a depth of at most about two feet. Of course, all these results depend upon the nature of the soil.

The thermometers nearest the surface are affected by changes in the weather, but these disturbances are not felt far down.

When a steady state of temperature is reached, r vanishes, and (3) becomes

$$kg'=0.$$

This gives

$$g=\text{constant},$$

which applies directly to the case of the earth regarded as a cooling body, and shows us that *in strata throughout which the conductivity does not vary, the temperature increases uniformly per unit of depth.* Of course, if the earth is regarded as a *cooling* body, the steady state of temperature is impossible; but its rate of cooling is so slow that the time-variations of temperature may be neglected.

Fourier's equations, when applied to past time as regards the earth or the sun, indicate a state of uniform high temperature throughout the mass—a state which could not have arisen by any process of conduction. This suggests the production of the heat by the gravitation of separate masses (§ 77).

260. *Conduction in Crystalline Bodies.*--Crystalline bodies possess, in general, unequal conducting powers in different directions. The conducting power is symmetrical with regard to three rectangular axes—called the principal axes of thermal conductivity.

If a single point-source of heat were placed in the interior of a crystal, the loci of constant temperature would be concentric ellipsoids surrounding that point; and Stokes has shown that the conductivities parallel to the axes of these ellipsoids are proportional to the squares of the axes. Sections of such an ellipsoid can be obtained by means of thin plane plates of the crystal cut in different directions from the substance. If a copper wire be passed through a small hole drilled through such a plate in the direction of the axis of the ellipsoid conjugate to the plane, and if this wire be heated by an electric current, the heat conducted away by the plate may be made to melt a thin coating of beeswax on the surface of the plate. The boundary between the melted and the unmelted wax is a section of the above ellipsoid.

C. H. Lees, using a device of O. Lodge, placed a crystalline plate between two portions of a brass bar with amalgamated ends, and determined the conductivity of the plate by measurements of the temperature gradients in the two portions of the bar whose conductivity was previously found by Forbes' method.

261. *Conduction in Liquids and Gases.*—The thermal conductivities of liquids (neglecting liquid metals) are small in comparison with those of solids; and those of gases are smaller still.

In experimental investigations on the subject, great care must be taken to avoid convection currents (see below), which would completely invalidate the results. The results might also be rendered inaccurate through transference of heat by radiation or through inter-diffusion of relatively hot and cold parts of the liquid.

Depretz applied heat at the surface of a liquid which was contained in a wooden vessel in order that there might not be appreciable conduction of heat through the walls. Guthrie used two equal hollow metal cones, whose horizontal flat bases were placed opposite each other. The liquid was retained by capillarity in the space between the cones. The lower cone was used as an air thermometer, and the upper cone was heated by a current of hot water. Some seconds elapsed after the hot water was turned on before the air thermometer registered any difference of temperature. This showed that radiation had no appreciable effect—a fact which Guthrie also proved otherwise. If the face of the upper cone was coated with a soluble aniline dye no appreciable diffusion took place during the experiment, so that transference of heat by diffusion was negligable.

In another method flat metal plates, having the liquid contained in the space between them, are used. The lower plate is cooled by ice, and the flow of heat is determined by the rate of cooling of the upper plate. In yet another method the liquid is enclosed in the narrow space between two vertical cylinders. The following numbers give conductivities in centimetre-gramme-minute units:

Benzine ...	0·0235	Water at 4°·1 ...	0·0745
Glycerine ...	0·0430	Mercury at 4°·5 ...	0·9250

These results, among others, are due to Weber and Lorberg.

Increase of temperature raises the conductivity of a liquid.

Magnus determined the relative conductivities of some gases by a method similar to that used by Depretz in his experiments on liquids. The more recent determinations of absolute conductivities are based upon observations of the cooling of a thermometer in the gas. It is found that, through a considerable range of pressure, the rate of cooling is constant. This shows that convection has ceased to have any appreciable effect. Kundt and Warburg found that, in air, this range was from 150 mm. to about 1 mm., while in hydrogen it was from 140 mm. to about 9 mm. The correction which has to be applied because of radiation is found by a separate experiment, in which the enclosure containing the gas is exhausted as far as possible. Kundt and Warburg give, as the relative values of the conductivities of carbonic acid, air, and hydrogen, the numbers 0·59, 1, and 7·1 respectively. Winkelmann gives 0·000052 as the absolute value for air in centimetre-gramme-second units.

In the case of gases, the conductivities can be calculated from the kinetic theory, and the results agree well with observation.

262. *Convection.*—Under gravity, all liquids and gases tend to arrange themselves in horizontal layers the densities of which decrease as their distances from the earth's surface increase. This condition may be entirely disturbed because of variations of temperature; for the consequent changes of density destroy the equilibrium, and currents are set up in the fluid so as to restore it. These are called 'convection-currents.'

We have already discussed a typical case in dealing with the maximum density point of water.

Very marked examples occur on a large scale in nature. The trade winds are due to the ascent of hot air currents in equatorial regions, while colder air blows in from the polar regions to take its place. [The north-easterly or south-easterly direction of these winds is due to the rotation of the earth.] A considerable part, at least, of ocean circulation, is also of the nature of convection. Again, whenever water evaporates, heat is absorbed to be evolved wherever condensation occurs. This is indeed one of the most fruitful sources of violent storms; for, if sufficient heat is developed, the consequent increase of temperature causes a rapid up-rush of air, and so creates a partial vacuum which gives rise to a violent inflow of the surrounding air. The air which comes from the south has (in the northern hemisphere) a greater eastward motion than that which comes from the north, and so a counter-clockwise vortex motion is produced. Thus the rotation of a cyclone is explained.

Practical applications of the principles of convection are seen in the usual methods of boiling water, of ventilation, etc.

A hot body, which is cooling in a gas or a liquid, loses heat by convection as well as by radiation. The law of convective cooling in a gas has been elaborately studied by Dulong and Petit. Their results are expressed by the formula

$$r = ap^b\theta^{1.223},$$

where r is the rate of cooling, a is a constant for a given gas and a given body, p is the pressure of the gas, b is a constant for any one gas, and θ is the excess of the temperature of the cooling body over that of the gas. The rate is independent of the nature of the surface of the body, but varies with its form and dimensions.

CHAPTER XXIII.

THERMODYNAMICS: HEAT AND WORK.

263. *Mechanical Equivalent of Heat. First Law of Thermodynamics.*—Heat, since it is a form of energy, may be transformed into mechanical work and into all other forms of energy; but the former transformation is the only one with which we are at present concerned.

Rumford was the first to make a determination of the amount of work which can be produced from a given amount of heat, *i.e.*, of the *mechanical* (or, more properly, the *dynamical*) *equivalent of heat*. As the result of his experiments, already alluded to, on the heat which was developed in the process of boring cannon, data are given by him from which the equivalent may be calculated; but the value is much too large.

In 1837, 1839, 1842, and 1843 respectively, Mohr, Séguin, Mayer, and Colding made suggestions regarding the relations between heat and work, which, although—by reason of deficiency of experimental data—leading to no accurate numerical result, are yet very noteworthy in the historical development of the subject and exhibit great scientific insight.

The most direct method of conducting such an investigation consists in spending, as Rumford did, a known amount of work in the production of heat by friction. This method was also used by Joule, who caused a falling weight to drive a vane rotating in the interior of a calorimeter which contained a known amount of water. The amount of heat developed was determined by means of the observed increase of the temperature of the water, due precaution being taken to correct for the loss of heat by radiation, and for the heat developed by friction between the parts of the apparatus. Various other methods were used by Joule, Hirn, Regnault, and others. For example, Joule proved experimentally that the heat developed by the sudden compression of air is practically equivalent to the work spent in compression (§ 268); and from this result, together with an accurate determination of the specific heats of air, he obtained the value of the mechanical equivalent. He also determined its value by means of the heat developed in a conductor by the passage of a current of electricity through it under given conditions (Chap. XXVI.). One of Hirn's series of experiments was made upon a heat-engine in actual use. In another series, he experimented upon the heat developed by percussion. The latter of these gave a good result; the former did not.

Joule finally gave the number 772 (in foot-pounds at the latitude of Manchester) as the amount of work necessary to raise the temperature of one pound of water by one degree Fahrenheit. In foot-pound centigrade-degree units the value is therefore 1390. In kilogramme-metres this becomes about 428; and, in ergs, the value is about forty-two millions. Rowland and Miculescu have recently got results in close agreement with Joule's.

The above experiments prove the law of conservation of energy in so far as heat and work are concerned. The statement of the law of conservation for these two forms of energy goes by the name of the FIRST LAW OF THERMODYNAMICS, which asserts that *when equal quantities of mechanical effect appear from purely thermal sources, or disappear in the production of thermal effects alone, equal quantities of heat disappear or are produced.*

264. *Carnot's Complete Cycle of Operations.*—Although, as is indicated in the previous section, we can determine, by experiment, the direct relation between given amounts of heat and work, we are not entitled to draw any conclusion regarding the relation between the heat which disappears and the work which appears in any given physical process, unless a certain condition be observed. The necessity for this condition (which has already been referred to in § 228) was pointed out by Sadi Carnot.

The condition is that the working substance must pass through a complete cycle of operations, *i.e.*, a cycle at the end of which the substance has returned to its original physical state. Heat may have been expended in the given cycle, and work may have been produced; but, unless the final state is the same as the initial state, we cannot say that the work and the heat are mutually equivalent. For example, carbonic acid gas is heated by compression, but the heat developed is not the equivalent of the work spent in compression, for work is done by the molecular forces during the process.

In order to be able to reason correctly upon the connection between heat and work, Carnot assumed the existence of a heat-engine which can never be realised in practice. But this does not render his results any the less valuable, for, in order that we may be able so to modify Carnot's results that they may apply to any special case, we only require to know in what way, and to what extent, the given engine differs in its action from Carnot's.

He assumed that his engine was furnished with a cylinder the sides and piston of which were absolutely impermeable to heat, while the bottom was a perfect conductor of heat. He assumed also the existence of two bodies, one hot and the other cold, the temperatures of which were kept constant. The former of these was to act as the source of heat; the latter was to act as the condenser. The working substance is supposed to be placed in the cylinder below the piston, and may be any substance whatsoever, with any properties whatsoever. But, for the sake of definiteness, we may assume that it acts as the steam in an ordinary engine would act.

Let us suppose that the cylinder, with its contents at the temperature of the cold body, is placed on a non-conductor of heat. The contents will retain their temperature (t_0, say) so long as the cylinder remains on the non-conductor, for the working substance is now surrounded on all sides by non-conductors. Let the volume and pressure of the substance be denoted by v_0 and p_0 respectively. [Remark here that the physical condition of a known mass of the substance is completely determinate when any two of its temperature, volume, and pressure are given.]

As the first operation of Carnot's cycle, the cylinder still remaining on the non-conductor, press down the piston until the temperature of the substance rises to that of the hot body (t_1, say) and the volume and pressure become v_1 and p_1 respectively.

As the second operation of the cycle, place the cylinder, with the condition of its contents unaltered, on the hot body, and let the substance slowly expand until its volume becomes (say) v_2, and the pressure becomes p_2 ($<p_1$). The expansion must occur so slowly that heat can flow into the substance so as to constantly prevent its temperature from being finitely different from t_1.

As the third operation place the cylinder and its contents—again without variation of condition—upon the non-conductor, and let the working substance expand until its temperature falls to that of the cold body. Let the pressure and the volume now be p_3 ($<p_2$) and v_3 respectively.

Lastly, place the cylinder upon the cold body and slowly press down the piston—the temperature of the contents remaining at t_0 until the volume again becomes v_0, and the pressure, therefore, again takes the value p_0.

This series of operations obviously satisfies Carnot's condition that the final condition of the working-substance shall be identical with its initial condition, *i.e.*, it forms a complete cycle.

Now in the second operation heat was taken from the hot body, and, in the fourth operation, heat was given to the cold body. Let these quantities be h_1 and h_0 respectively.

Also, in the first and fourth operations, work was done upon the contents in diminishing their volume; and, in the second and third operations, the contents, in their expansion, did work against the external pressure. Let these quantities be w_0 and w_1 respectively. In the fourth and first operations the volume decreased from v_3 to v_1, and the temperature was either equal to t_0 or rising from t_0 to t_1; in the second and third operations the volume increased from v_1 to v_3, and the temperature was either equal to t_1 or falling from t_1 to t_0. In the former pair the temperature was therefore on the whole lower than it was in the latter pair. Consequently the pressure was higher when work was being performed by the substance than when it was being expended upon it, and therefore $w_1 - w_0$ is positive. And, the cycle being complete, we can write

$$J(h_1 - h_0) = w_1 - w_0,$$

where J is the multiplier (the mechanical equivalent, sometimes called Joule's equivalent—hence the letter J) required to change the heat units into dynamical units. This equation is the analytical expression of the First Law of Thermodynamics.

265. *Carnot's Reversible Cycle.*—Besides the idea of a complete cycle of operations, Carnot introduced the equally important and fruitful idea of a *Reversible Cycle.* This is a cycle which can be performed in the exact reverse order; and Carnot's cycle can be so performed.

First. Begin with the cylinder on the non-conductor, the temperature, volume, and pressure of its contents being t_1, v_1, and p_1 respectively. Let the substance expand until these quantities become t_0, v_0, and p_0.

Second. Place the cylinder on the cold body and let the expansion proceed until the pressure and volume become p_3 and v_3 respectively, the temperature being still t_0.

Third. Place the cylinder on the non-conductor, and push down the piston until the temperature rises to t_1, the pressure and volume becoming p_2 and v_2 respectively.

Fourth. Place the cylinder on the hot body and compress the contents until the initial conditions are again attained.

Now in the first and second of these reverse operations work was done by the substance, while the temperature had either the low value t_0 or was falling from t_1 to t_0; and, in the third and fourth operations, work was done upon the substance, while the temperature remained at the high value t_1, or rose from t_0 to t_1. On the whole, therefore, the temperature—and consequently the pressure—had a higher value when work was done upon the substance than when work was performed by it. But heat was absorbed from the cold body in the second operation, and was given to the hot body in the fourth. In the reverse cycle, therefore, heat has been pumped up from the condenser to the source of high-temperature heat, but work has been expended in the process. The complete action is also represented by the equation of last section.

Carnot reasoned upon the supposition that heat was material, and so believed that the quantity of heat which was absorbed from the hot body was given to the cold body in the direct process, and that the quantity which was absorbed from the cold body was given to the hot body in the reverse process. He supposed that the heat did work in the direct process merely in being let down from a source at high temperature to a sink at low temperature, just as water does work in falling from a high to a low level.

The interpretation of his result, on the principle of conservation of energy, is simply that the excess of the heat absorbed over that emitted is directly transformed into its equivalent in work; and that, in the reverse operation, the heat-equivalent of the work expended, together with the heat absorbed from the cold body, is equal to the heat given to the hot body.

266. *Reversibility the Test of Perfection. Second Law of Thermodynamics.*—We shall now discuss one of the many important results which can be deduced from Carnot's principles.

Let an engine be reversible in the sense that all its physical and mechanical actions can be performed in the exact reverse order. Such an engine is perfect in the sense that it is as perfect as any engine, working under the same conditions, can be, *i.e.*, the reversible engine, taking in a quantity h_1 of heat at temperature t_1, and working with its condenser at the temperature t_0, will perform as much work as will any other engine, working through the same range of temperature, and also taking in the quantity of heat h_1 at the temperature t_1.

Carnot proved this statement by showing that, if it were not true, the perpetual motion would result. Let A denote the reversible engine, and let B denote the (supposed) more perfect engine. Make B work directly between the temperatures t_1 and t_0, taking in a quantity of heat h_1, giving out a quantity h_0, and performing an amount of work W, > w, the quantity of work which the reversible engine would produce under the same conditions. Make A work backwards between the same sources at the same high and low temperatures. A quantity of work w is all that has to be expended in order to make it take in the amount h_0 of heat at the temperature t_0 and restore the amount h_1 to the source at the high temperature. Hence an excess of work $W - w$ is gained in the double process, while, on the whole, no heat has been transferred either way. But this means the perpetual production of work from nothing, which is impossible. Hence the reversible engine is perfect.

To adapt this reasoning of Carnot to the modern ideas of energy, we have only to argue thus: Work is performed in the double cycle, while no heat is on the whole taken from the source, therefore B must give less heat to the cold body than A takes from it, and so the double engine can only work by giving to the hot body heat which it has taken from a colder body, the temperature of which it constantly lowers; but this is in opposition to all known facts, and the denial of its possibility may be safely taken as axiomatic.

This adaptation is due to Lord Kelvin, who re-introduced Carnot's work to the scientific world when it had been long disregarded, and who applied his principles to the deduction of thermodynamical results of the highest importance.

The statement that an engine which is reversible, in the sense that all its physical and mechanical actions are capable of exact reversal, converts, under given conditions, the greatest possible fraction of the heat which is supplied to it into useful work is known as the SECOND LAW OF THERMODYNAMICS. We shall find in § 270 that its analytical expression is $\Sigma(h/t) = o$.

267. *Absolute Temperature.*—The *efficiency* of a heat-engine is the ratio of the quantity of heat which it utilises in the form of mechanical work to the total quantity of heat which is supplied to it. In the notation used above it is

$$\frac{h_1 - h_0}{h_1}.$$

The value of this fraction is a maximum, under any given conditions, when a reversible engine is used. And *all* reversible engines, whatever may be the nature and properties of the working-substance, are equally perfect (*i.e.*, they possess the same efficiency) when they work through the same range of temperature. This enables us, as Kelvin pointed out, to obtain an *absolute* measure of temperature—absolute in the sense that it does not depend upon the properties of any particular substance.

The definition which Kelvin finally adopted (see § 269) was framed so as to make the absolute scale coincide as nearly as possible with the scale of the air-thermometer. It is this:

'*The temperatures of two bodies are proportional to the quantities of heat respectively taken in and given out in localities at one temperature and at the other, respectively, by a material system subjected to a complete cycle of perfectly reversible thermodynamic operations, and not allowed to part with or take in heat at any other temperature; or, the absolute values of two temperatures are to one another in the proportion of the heat taken in to the heat rejected in a perfect thermodynamic engine, working with a source and refrigerator at the higher and lower of the temperatures respectively.*'

Expressed in symbols, this gives

$$\frac{t_1}{t_0}=\frac{h_1}{h_0},$$

where t_1 and t_0 are the absolute temperatures of the source and the condenser, respectively.

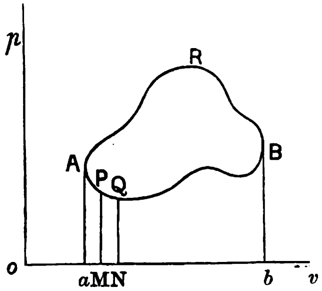


FIG. 140.

268. *The Indicator Diagram.*—The indicator diagram was introduced by Watt for purely practical purposes. It is, nevertheless, as we shall shortly see, susceptible of numerous important applications in pure science.

The diagram exhibits the relation between the pressure and the volume of the working substance used in any heat-engine; and hence it shows the work done in a complete stroke of the engine.

Let the curve APQBR (Fig. 140) represent the relation between the pressure and the volume at all stages of the stroke—pressure and volume being respectively measured along the rectangular axes

op and ov. Let the co-ordinates of the point P be $PM = p$, $OM = v$; and let those of Q be $QN = p'$, $ON = v'$. When P and Q are indefinitely near each other, the product $(p+p')(v'-v)$ represents twice the area PMNQ. But, neglecting small quantities of the second order, half this product may be written $p(v'-v)$. This latter product therefore represents the elementary area PMNQ.

Now the work done by a force f, acting through a distance s, is (§ 43) fs; and the whole force which acts upon the piston in the cylinder of the engine is pa, where a is the area of the piston, and p is the pressure per unit area, so that pas represents the work done when the piston moves through the small distance s under the action of the pressure pa. But as is the small change of volume of the contents of the cylinder which is produced by that pressure; and therefore the work done is represented by the area PMNQ.

Let A be the point at which the volume has its smallest value, v_0, and let B be the point of maximum volume, v_1. As the volume expands from v_0 to v_1, the state of the substance being represented by a point which moves along the path APQB, the work performed is represented by the area A*ab*BQPA. Similarly, when the volume is diminished from v_1 to v_0, the point moving from B to A along the path BRA, the work expended in producing compression is represented by the area BRA*ab*B. The difference of these two areas, viz., the curvilinear area APQBRA, therefore represents the work which is expended, on the whole, during a complete stroke of the engine. Consequently, when the path is described in the positive direction, work is expended on the whole. This case corresponds to the reverse working of a reversible engine.

When the closed path is described in the negative direction, work is performed on the whole by the engine to an extent which is represented by the total area enclosed by the path. [The actual path might consist of a number of closed loops. In this case, each loop is to be considered separately, and the sum of the areas—each with the proper sign attached, according as it is described positively or negatively—is to be taken in order to estimate the total amount of work performed.]

269. *Applications of the Indicator Diagram.*—We shall now consider the application of the diagram to the discussion of the working of Carnot's engine and the determination of the definition of absolute temperature.

For this purpose we must give the closed curve a special form. In the second and fourth direct operations of Carnot's cycle (§ 264), the substance was maintained at constant temperature; in the first and third, no heat, *as such*, was allowed to pass out of it or to enter it. If, therefore, the points D, A, B, C (Fig. 141) represent respectively the values (p_0, v_0; p_1, v_1; p_2, v_2; p_3, v_3) of the pressure and volume of the substance at the commencement of the first, second, third, and fourth operations respectively, all points on the line DA (which represents the varying state of the substance as it passes from the state (p_0, v_0, t_0) to the state (p_1, v_1, t_1) are characterised by the condition that no heat, as such, enters or leaves the working substance; all points on the line AB represent states in which heat is absorbed during the passage from the condition A to the condition B in order that the temperature may retain the value t_1; all points on the line BC represent states in which, again, no heat enters or leaves the substance; and all points on the line CD represent states in which the temperature of the substance is, by disengagement of heat, maintained at the value t_0.

Lines such as AB or CD are therefore called *isothermals*, while lines such as DA or BC are called *adiabatics*. The latter name,

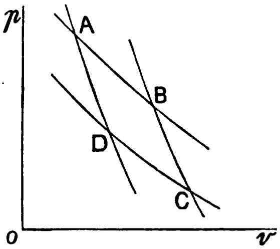


FIG. 141.

which is due to Rankine, simply implies that no heat is absorbed or emitted.

But although heat, *as such*, neither enters nor leaves the substance in the adiabatic condition, mechanical work is performed or expended; and so, *by transformation*, the amount of heat contained in the substance is actually varying.

If the amounts of heat contained in the substance, in the states A and D, were identical, and if the amounts in the states B and C were identical, the amounts of heat which leave the substance in the processes represented by AB and DC would necessarily be equal. But we know that these quantities are (by our definition) in the ratio

$$\frac{h_1}{h_0} = \frac{t_1}{t_0}, \quad \dots\dots \quad (1),$$

t_1 and t_0 being absolute temperatures.

Now ABCD represents the work which is performed in the direct cycle. It, therefore, by the principle of conservation of energy, represents the quantity of heat $h_1 - h_0$. And, if we arrange matters so that $h_1 - h_0 = 1$, equation (1) shows that $t_1 - t_0$ must be equal to unity if we make $h_1 = t_1$ and $h_0 = t_0$.

Let us therefore intersect the diagram (Fig. 142) by a series of isothermals $A_1A_2A_3\ldots$, $B_1B_2B_3\ldots$, etc., and by a series of adiabatics $A_1B_1C_1\ldots$, $A_2B_2C_2\ldots$, etc.; and let the temperature corresponding to $B_1B_2B_3\ldots$ be one degree below that corresponding to $A_1A_2A_3\ldots$, while the adiabatics are so arranged that the magnitude of each area, such as $A_1A_2B_2B_1$, is unity. If we could continue this construction down to the absolute zero of temperature, each area included between any isothermal, the isothermal of absolute zero, and any two consecutive adiabatics, would be numerically equal to the quantity of heat taken in during the expansion between the two adiabatics and also to the temperature indicated by the higher isothermal.

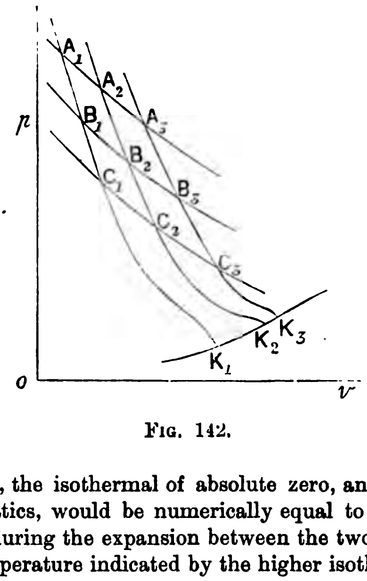


FIG. 142.

The experimental data necessary to the correct completion of the diagram are wanting, but its correct completion is not necessary to our present purpose. The following method of completing it is due to Maxwell:

Let us suppose that $C_1C_2C_3\ldots$ is the lowest isothermal whose form is correctly known. Draw any line $K_1K_2K_3\ldots$ to represent the isothermal of absolute zero, and complete the adiabatics in such a way that each of the areas $C_1K_1K_2C_2$, $C_2K_2K_3C_3$, etc., is numerically equal to the temperature to which the line $C_1C_2C_3$... corresponds. In order to determine the position of the absolute zero, it is only necessary to find the ratio of any two areas such as $A_1K_1K_2A_2$ and $C_1K_1K_2C_2$.

We have already seen that the work done by a perfect engine does not depend on the nature of the working substance. Let us therefore suppose that the working substance is the hypothetical substance dealt with in the kinetic theory of gases. Since this substance is a monatomic gas, whose atoms are supposed to be smooth hard spheres, and to exert no interatomic forces, all internal energy supplied to it takes the form of kinetic energy of atomic motion. If work be expended in compressing it, an exact equivalent of internal heat energy appears; and the total amount of internal kinetic energy which it contains at a given instant is measured by the product of its pressure and volume. If, therefore, we *assume* that we are to measure the temperature by this product, we shall find that we have a substance which, when used in a constant-volume or constant-pressure gas thermometer, will give indications of temperature agreeing with those determined by means of Kelvin's definition of absolute temperature. For, suppose that the engine works in a cycle such as that represented by $A_1A_2B_2B_1A_1$ in the above diagram. In the expansion from A_1 to A_2 at the constant temperature θ, say, the internal energy remains constant, and so the work which is done is the equivalent of the heat h, say, which is taken in. In the expansion from A_2 to B_2, no heat is taken in or given out *as such*, though internal energy is spent in the performance of external work. In the compression from B_2 to B_1, the internal energy again remains constant, and an amount of heat, h', is given out, which is the exact equivalent of the work which is expended. Finally, in the compression from B_1 to A_1, no heat, as such, passes in or out. But, in the second and fourth operations respectively, the loss and gain of internal energy are equal, for each of these quantities is measured by the difference of the products of pressure and volume corresponding to the two isothermals. Now the work done in the first operation is represented by the area included by the isothermal θ, the axis of volume, and the lines of constant volume passing through the points A_1 and A_2; and so on with regard to the other similar cycles, $A_2A_3B_3B_2A_2$, etc. But the work which is done in the first operation of each of these cycles is also the equivalent of h, while that expended in the third is the equivalent of h'. Hence the ratio of h to h' is equal to the ratio of the areas included by the axes of volume and of pressure and the isothermals θ and θ' respectively. And the ratio of these areas is the ratio of the products of pressure and volume for each isothermal respectively; for, at a given volume, the ratio of the ordinates of the two isothermals is the ratio of the corresponding pressures. Therefore, with such a working substance, the absolute temperature as above defined *must* be proportional to the product of the pressure and volume; and the temperature registered by a constant-volume or constant-pressure gas thermometer filled with the given substance is also proportional to it by our fundamental assumption.

It follows that any gas, for which the well-known law $pv = \mathrm{R}t$ holds, would, if employed in a gas thermometer, give absolute temperatures; and the scales of the so-called permanent gas thermometers will not differ much from the absolute scale. Since an absolutely ideal perfect gas is unknown, the true absolute scale must be determined by observations of the manner in which some gas, such as air, differs from a perfect gas (see Chap. XXXI.).

270. *Applications of the Indicator Diagram. Entropy.*—The one set of lines with which we have intersected the diagram are characterised by constancy of temperature. The other set of lines are characterised by the condition that no heat shall be transferred from the working substance to its surroundings, or from the surroundings to it. But we have seen that heat does pass from, or into, the substance by transformation. We have still to inquire, therefore, What quantity remains constant during adiabatic expansion or contraction?

Equation (1) of last section enables us to answer the inquiry. It gives

$$\frac{h_1}{t_1} = \frac{h_0}{t_0},$$

whatever values t_1 and t_0 may have; that is to say, when the substance passes isothermally from any one definite adiabatic state to any other definite adiabatic state, the quantity of heat which is absorbed, or disengaged, bears a definite ratio to the temperature at which the absorption, or disengagement, takes place.

This constant quantity h/t may therefore be regarded as the amount by which some quantity, ϕ, which is characteristic of the substance in the adiabatic state, changes in passing from the one adiabatic condition to the other at constant temperature t. This suggests the extension of the meaning of the quantity h to signify the total heat contained in the substance, so that we may define ϕ by the equation

$$\phi = \frac{H}{t},$$

where H represents the total heat.

The quantity ϕ was first called, by Rankine, the *Thermodynamic Function*. Clausius called it the *Entropy*, and this name has been generally adopted.

If one substance parts with an amount of heat, h, at temperature t, to another substance at temperature t_0, the entropy of the former substance decreases by the amount h/t_1, and that of the latter increases by the amount h/t_0. The total gain of entropy by the system is therefore

$$h\left(\frac{1}{t_0} - \frac{1}{t_1}\right).$$

Similarly, the total gain of entropy of a system consisting of a number of sources and sinks may be represented by

$$\Sigma\left(\frac{h_0}{t_0} - \frac{h_1}{t_1}\right).$$

Deleting the suffixes, we may denote this simply as

$$\Sigma\left(\frac{h}{t}\right),$$

and our definition of absolute temperature shows that *this vanishes when a perfect engine is used.* When any other engine is used, heat is always lost by conduction or otherwise, so that the heat which is given to the condenser (or is otherwise wasted) is greater than h_0. Hence, in all actual cases of the transformation of heat into work we must write

$$\Sigma\left(\frac{kh_0}{t_0} - \frac{h_1}{t_1}\right),$$

instead of

$$\Sigma\left(\frac{h_0}{t_0} - \frac{h_1}{t_1}\right),$$

where k is greater than unity, as the proper expression for the gain of entropy. This is equal to

$$\Sigma\frac{h_1}{t_1}(k-1),$$

which is necessarily *positive,* since k is greater than unity; and so we prove Clausius' theorem that *the entropy of the universe tends to a maximum.*

The quantity $\Sigma h_1(k-1)$ evidently represents the energy which is unnecessarily wasted.

The quantity of heat which is absorbed in the second operation of Carnot's direct cycle is $t_1(\phi_1-\phi_0)$, and the quantity which is given out in the fourth operation is $t_0(\phi_1-\phi_0)$. The difference of those quantities is (t_1-t_0) $(\phi_1-\phi_0)$, which therefore represents the work done in the complete cycle.

271. *Applications of the Indicator Diagram. Total, Available, and Dissipated Energy.* — We have no means of determining experimentally the total amount of energy in a given system. But, in any practical case, we only require to determine the *change* of energy which takes place in the given operations.

The indicator diagram enables us to represent the *total* amount of energy in a way which, though obviously incorrect, leads to a correct representation of the *change* of the total energy of the system which is produced by a given change in its physical condition.

Let the points A and B (Fig. 143) represent respectively the initial and the final conditions of the system, and let the path AB represent the series of changes by which the final condition was arrived at. [The diagram is constructed so as to exhibit the case in which there is both disengagement of heat and performance of mechanical work.

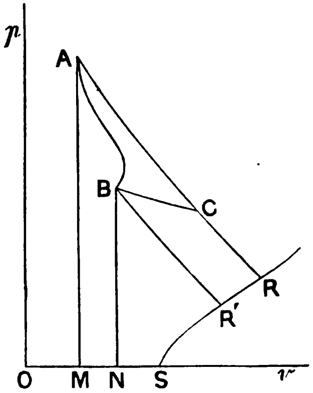


FIG. 143.

Let RR′ represent the (arbitrary) isothermal of absolute zero, and let it be continued so as to cut the axis of volume in the point S.

As the point which traces out the diagram moves from A to B, external work, which is represented by the area ABNMA, is performed, and at the same time (§ 269), an amount of heat is disengaged which is represented by the area ABR′RA. The whole area AMNBR′RA, therefore, represents the total loss of energy which the working substance has sustained. We may, therefore, as Maxwell suggested, regard the areas AMSRA and BNSR′B, respectively, as representing the total amounts of energy which are contained in the working substance in the conditions indicated by A and B respectively.

It is specially to be noticed that the amount of energy which is lost in proceeding from A to B is totally independent of the path AB, and depends only on the initial and final conditions of the substance.

Let us assume now, as a special case, that BC represents the lowest available temperature. We may then cause the substance to pass from the state A to the state B along the path ACB. The work performed will then be represented by the area ACBNMA, which area therefore represents the total amount of energy which is available, under the given conditions, for the performance of mechanical work. Similarly, BCRR′B represents the heat which is necessarily given to the condenser, *i.e.*, the amount of energy which is necessarily dissipated so far as the performance of work by the given system is concerned. The energy which is unnecessarily dissipated is represented by the area ACB.

272. *Thermodynamic Motivity.*—We have already seen (§ 267) that the efficiency of a heat engine is represented by

$$\frac{h_1-h_0}{h_1}$$

In a perfect engine this becomes

$$h_1\frac{t_1-t_0}{t_1}$$

Kelvin has called the total energy which could be made available for mechanical work by a perfect engine under given conditions, the *Thermodynamic Motivity* of the system. If we have a medium external to the given system, which may be used as a condenser, the motivity is the whole amount of work which can be obtained by the perfect engine in reducing the temperature of the system to that of the external medium. If the engine works so as to equalise the temperatures of the various parts of the system, the motivity is the whole amount of work which can be so obtained.

If t_0 be the final temperature, while h_1 is a quantity of heat taken from a body at temperature t_1, the motivity, so far as this quantity is concerned, is represented by the above expression for the efficiency of a perfect engine; and, to obtain the total motivity, we must sum all such quantities.

The total energy, e, in any system is equal to the sum of the motivity, m, and the dissipated energy of that system. If ϕ is the entropy, θ the temperature, and J the mechanical equivalent of heat, the dissipated energy is $J\theta\phi$. Hence, if the system passes from a state indicated by the suffix 1 to a state indicated by the suffix 2, we get

$$m_1 - m_2 = e_1 - e_2 - J(\theta_1\phi_1 - \theta_2\phi_2).$$

No change can take place of *itself* in the system unless thereby the motivity be decreased—that is, unless $m_1 - m_2$ be positive. But $m_1 - m_2$ may be positive, although $e_1 - e_2$ is negative, provided that $\theta_2\phi_2$ is sufficiently greater than $\theta_1\phi_1$. Hence we see that a given chemical action may take place of itself with absorption of heat, provided that a sufficient amount of energy be dissipated in the process. (See § 254.)

CHAPTER XXIV.

ELECTROSTATICS.

273. *Electrification by Friction.*—A rod of glass, when rubbed with leather coated with a paste of zinc amalgam, acquires the property of attracting surrounding bodies. Light bodies, such as pieces of paper, can even be raised up by it against the attraction of the earth. When in this state, the glass is said to be *electrified* —or it is said that *electricity* has been developed upon the glass.

If the glass had been rubbed with cat's fur or with any one of several other substances, similar effects would have ensued; but the extent to which electrification is developed depends upon the nature of the substance which is used as a rubber.

The glass may be replaced by sealing wax, resin, ebonite, etc., and the phenomena will still be exhibited to a greater or less extent.

In all cases it is practically necessary for success that the substances be kept warm and dry.

274. *Conductors and Non-Conductors.*—If, instead of a glass rod or a rod of sealing-wax, we take a metallic rod, no electrical effects are in general observable; and many other substances also are incapable (unless special means are adopted, § 292) of being electrified by friction.

We are thus led to divide all substances into two classes according as they are or are not electrifiable by friction in the usual way. Those of the former class are called *Insulators*, *Dielectrics*, or *Non-conductors*; those of the latter class are called *Conductors.* The latter terms are applied because it is found that all substances which cannot usually be electrified by friction have the power of allowing electricity to flow along them, while the other class of substances practically prevent such flow.

275. *Fundamental Phenomena presented by Electrified Bodies.* —The substances which are attracted by an electrified body are not

necessarily non-conductors. In order to investigate the subject further, we shall suppose that a pith ball (which is a conductor, and is at the same time very light, so that the effects to be observed are easily seen) is the body to be attracted, and we shall suppose it to be insulated by being suspended from a dry glass rod by means of a dry silk thread.

If a glass rod, electrified by friction with zinc amalgam, be brought into the neighbourhood of the pith-ball, the ball will be drawn towards it; and this will also take place when sealing-wax, electrified by friction with flannel, is presented.

Now let the glass rod be brought so near that the ball comes into contact with it. Immediately after contact the ball is violently repelled by the glass; but if the ball be touched with the hand attraction will again occur; and the same phenomena will happen when electrified sealing-wax, or any other electrified body, is used. Still, though all electrified bodies produce this effect, a slight modification of the experiment will bring to view a profound difference in the nature of the electrification of different substances.

Instead of touching the pith-ball when it is repelled by the glass, let the electrified sealing-wax be presented to it. Strong attraction becomes apparent. Similarly the electrified glass will attract the ball when the sealing-wax repels it. And all substances which can be electrified by friction can be classified according as they act in this respect like the glass or like the sealing-wax.

The following substances, in the order given—cat's fur, flannel, glass, cotton, silk, metals, sealing-wax, resin, sulphur—are such that any one acts in one way when rubbed by those following it in the list, and in the opposite way when rubbed by those preceding it in the list. The series may be greatly extended, but the substances given are typical.

276. *Positive and Negative Electricity.*—In order to explain these phenomena we may make the following assumptions: 1st. There are two 'kinds' of electricity; 2nd. Like kinds repel each other, unlike kinds attract each other; 3rd. The attraction and repulsion diminish as the distance increases; 4th. An unelectrified body may be looked upon as a body which contains equal amounts of both kinds of electricity, which can be separated, to a greater or less extent, by means of the action of electrified bodies.

Let us distinguish the electricities developed on glass, by friction with silk, and on sealing-wax, by friction with flannel, as *positive* and *negative* respectively. When the positively electrified glass rod is brought near to the unelectrified pith-ball, we assert, in terms of our hypothesis, that the neutral electricities in the ball are separated, negative electricity coming to the side near the glass, positive electricity being repelled to the far side. The attraction between the unlike kinds is stronger than the repulsion between the like kinds, for the former are at a less distance apart. Hence the pith-ball moves towards the rod, for the electricity is confined to it and so cannot further alter its distance from the electricity of the rod unless the ball moves. If contact takes place, the negative electricity, which has been *induced* (as the phrase is, see § 278) in the ball, unites with a portion of the electricity of the rod, so that the ball is now charged with positive electricity, and is therefore repelled from the rod until it loses its charge (say, by repulsion to the ground when the observer touches the ball).

The same reasoning, with the interchange of the words positive and negative, applies when the sealing-wax is used instead of the glass.

Finally, the ball which has touched the glass rod is positively electrified, and is, therefore, attracted by the sealing-wax; and the ball which has touched the sealing-wax is negatively electrified, and so is attracted by the glass rod. All the phenomena are thus explained by means of our assumptions.

In the fourth assumption it was stated that an unelectrified body contains equal quantities of both kinds of electricity. In accordance with this assumption, it may be proved, by the methods to be shortly described, that the rubber which is used to produce electricity by friction becomes electrified to exactly the same extent as the rod which is rubbed, but with the opposite kind of electricity to that which is developed on the rod.

At one time it was customary to speak of electricity similar to that usually developed on glass by friction with silk as 'vitreous,' and of electricity similar to that which is produced on sealing-wax and other resins by friction with flannel as 'resinous,' electricity. The mere fact that the so-called resinous electricity may be obtained from glass is sufficient proof of the undesirability of this classification. The terms *positive* and *negative*, as we have employed them above, are much preferable, for the words imply nothing but a distinction in kind.

The phrase 'kind of electricity' is very apt to be misleading. We do not yet know what electricity is. One would never dream of saying that the resultant positive and negative forces which constitute a stress are essentially different from each other, or that left-handed (positive) rotation is intrinsically different from right-handed (negative) rotation. Yet equal and opposite forces, and equal and opposite rotations, annul each other's effects. The terms 'positive electricity' and 'negative electricity' are merely adopted in order to enable us to consistently and concisely describe and (so far) explain certain phenomena.

The use of the old expression 'electric fluid' is to be carefully avoided. Franklin in 1749 introduced the one-fluid theory. He assumed that the parts of the electrical fluid repelled one another, and that they were attracted by ordinary matter. Under usual conditions, ordinary matter was filled with the fluid. A defect of fluid constituted negative electrification; an excess of fluid, when present, surrounded the matter like an atmosphere, and constituted positive electrification. The two-fluid theory was originated in 1759 by Symmer. Each fluid was supposed to be attracted by the other and by ordinary matter, and repelled its own parts. The defects of these theories will appear subsequently.

277. *The Gold-leaf Electroscope.*—An *electroscope* is an instrument which is used to indicate the existence of electrification. If, in addition, the instrument measures the magnitude of the electrification, it is called an *electrometer*.

The gold-leaf electroscope is one of the most delicate of all electroscopes. It consists of two pieces of gold-leaf *a*, *a* (Fig. 144), which are connected, by means of a metal rod, to a metal head *h*. The rod passes through the top of a glass vessel in the manner indicated in the diagram. The glass vessel is open at the bottom, and contains a wire cage which surrounds the gold leaves. The cage can be placed in connection with the ground by means of the metallic connection *b*. The use of this cage will appear afterwards (§ 284); in the meantime we are merely concerned with the manner of using the instrument and the nature of its indications.

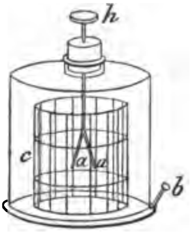


Fig. 144.

If a positively electrified body be brought into the neighbourhood of the head *h*, negative electricity is drawn towards the head, and positive electricity is repelled into the leaves, which diverge, since they are similarly electrified. The closer the body is brought to the head, the more widely do the leaves diverge; and, when the body is withdrawn, they collapse.

If, while the leaves are still diverging because of the presence of the electrified body, the head *h* be momentarily touched by the hand, instant collapse of the leaves will ensue (for the positive electricity escapes from the leaves through the hand to the ground), and the state of collapse will continue so long as the electrified body is not withdrawn. But, when the body is withdrawn from the neighbourhood of the head, the leaves once more diverge; for the negative electricity which was drawn to the head of the instrument now spreads in part through the metal rod into the leaves. The latter therefore are diverging with negative electricity.

In this condition the instrument can be used to indicate the nature of the electrification of any body which is brought into the neighbourhood of the head *h*. If the body be negatively electrified, more negative electricity will be repelled into the leaves which will therefore diverge more. If it be positively electrified (or be unelectrified), the negative electricity is drawn from the leaves, which then collapse; and if the positive electrification be sufficiently strong, some of the neutral electricity in the rod will be separated, and the leaves will diverge because of being positively electrified.

The interchange of the words positive and negative in the above reasoning will enable it to apply to the case in which a negatively electrified body is originally brought near to the head of the instrument.

It is obvious that these experiments are, in large part, merely a modified repetition of those which were discussed in § 275.

278. *Electric Charge and Quantity. Charging by Induction.* —In § 276 we have spoken of the electricity which is *induced* upon a conducting body because of the presence of another electrified body. So long as the bodies are not brought too near each other, the total amount of induced electrification is zero, a certain amount being drawn to one side of the conducting body, while an equal amount of the opposite kind is repelled to the other side. But whenever contact occurs, the attracted electricity unites with some of the electricity in the inducing body; and so the conductor is electrified with the same kind of electricity as that which the inducing body possesses. It is then said to have received a *charge* of electricity by subdivision of the charge which originally existed in the inducing body alone. The total effect is the same as if the inducing body had given some of its electricity to the conductor,

and it is usual to say that it *has* done so; for we cannot distinguish one amount of electricity from any other equal amount.

In the process of electrification by subdivision of a charge, the one body loses a certain amount of electricity, while the other gains an equal amount. This can be proved by means of measurements of the forces of attraction or repulsion which they exert upon an electrified body the electrification of which does not alter. In fact we can electrically *weigh* out equal amounts of electricity, just as we can gravitationally weigh out equal amounts of matter (see § 280). We are therefore justified in speaking of electricity as a thing which can be doled out in measurable quantities, and in using the word 'charge' in this sense of quantity of electricity.

Suppose, now, that we have a charged body—charged positively, let us say. It is possible by its means to charge other bodies, either positively or negatively, to any desired extent—and that without any reduction of its own charge.

Let A (Fig. 145) be the positively charged body, and let B and C represent two of the other bodies, of which B is to be negatively charged while C is to be positively charged. Each of the three bodies being well insulated from other conductors, place B and C in contact in some such position relatively to A as is indicated in the figure. Then let B and C be separated: B will be charged negatively, while C will have a positive charge. Greater effects would be produced, if necessary, by placing B and C at a considerable distance apart, and joining them by a thin conducting wire; for the effect of the charge in A is largely counteracted by the mutual attraction between the positive and negative electricities in B and C; and this mutual attraction is diminished as the distance between B and C increases. In practice, it is convenient to let C be the earth, and to place B in connection with it by means of a metallic wire or other conductor. In this case the repelled positive electricity in C is practically at an infinite distance. We may then use the body B, instead of A, if we wish to charge any other body negatively.

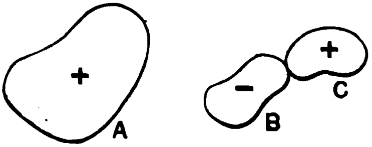


FIG. 145.

This process, in which the inducing body does not lose any of its charge, is called charging by *induction*.

The induced charge is, except in one special case (§ 284), less than the inducing charge.

279. *Continued Production of Electricity. The Electrophorus.*—In last section we saw how it is possible to obtain a positive or a negative charge at will by means of a single insulated charged body and two conductors. The instrument, based on this principle, which is generally used for the purpose, is the *electrophorus*. It consists of a flat, circular cake of resin, contained in a shallow metal vessel *ab* (Fig. 146). The resin is slightly warmed, and is electrified negatively by friction with cat's fur. A metal disc *cd*, insulated by means of a glass handle, is used instead of the conductor B of last section, while the earth takes the place of the conductor C.

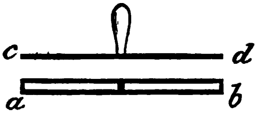


Fig. 146.

As the metal disc is brought near to the electrified resin, positive electricity is induced on its near side and negative electricity is repelled to its far side; and the more closely the disc is approached to the resin, the greater is the resultant separation of electricity.

When the disc is laid upon the resin, contact is made between it and the ground by means of a metal pin which passes through the centre of the cake of resin and is connected with the metal vessel enclosing it, and therefore with the ground upon which the vessel rests. The negative electricity escapes to the ground, and the disc is left charged with positive electricity, the charge being practically equal to that on the resin.

The disc may now be lifted away, and positive electricity can be communicated by contact from it to any conductor; and the process may be repeated from the beginning.

Two points in this explanation may present some difficulty. It may appear that the negative electricity of the resin should be destroyed whenever the disc is placed upon the surface of the cake; for the induced positive electricity in the disc would combine with it, and then the remaining negative charge in the disc would pass the ground through the metallic connection. This would really happen if the two surfaces came in contact throughout their whole extent; but, because of inequalities, they only touch over a comparatively small area. Again, the negative electricity in the disc is repelled by the electricity on the resin. How, then, can it pass to the ground by a connection which passes through the resin? We cannot be content with the reply that the disc and the earth are then both parts of one conductor, so that a road is opened up by which the electricity can get to a greater distance from that which repels it. This statement seems very like a statement to the effect that a stone would roll a short distance up one side of a hill in order that it might get a longer roll down the other side. Consider the apparatus before the disc is brought near it. The metal pin, being connected with the ground, is charged with positive electricity by the inductive action of the negative charge on the resin. If we were to draw out the head of the pin into a thin metal sheet, closely covering, but not touching, the surface of the resin, the positive charge would spread over this sheet in the same way as does the final positive charge on the metal disc when it is placed in contact with the pin, and might entirely prevent the inductive action of the negative charge on the resin. And the same action occurs, though to a less extent, when the head of the pin is not so enlarged. It is quite true that, when the disc is not too near the resin, the induced positive charge is entirely on its near side, while the induced negative charge is entirely on its far side. But when the disc is brought closer to the resin, the negative charge is drawn more and more to the centre of the near side of the disc by the action of the positive electricity in the pin, while the positive charge on the disc is drawn more to the outer parts of the near side by the action of the electrified resin. Finally, the positive charge in the pin unites with the negative charge in the disc. The earth is consequently left with an equal negative charge, and the effect cannot be distinguished from that which would have ensued upon an actual passage of the negative electricity of the disc to the ground. We have no means of distinguishing between the two cases, and therefore we are justified in saying that the electricity does pass from the disc to the ground.

In some cases, connection is made between the disc and the rim of the metal vessel which contains the resin, instead of between the disc and a central pin.

The production of electricity by means of the electrophorus becomes more and more continuous the more rapidly the various motions of the disc are performed. The principle of all machines used for the production of a statical charge is the same as that of the electrophorus, but they are so constructed as to give a strictly continuous production.

280. *Law of Electric Attraction and Repulsion.*—We have hitherto explained the various facts which have come under our consideration by the assumption of attractive or repulsive force, which diminishes in intensity as the distance between the attracting or repelling quantities increases; and we found that the assumption enabled us to give a consistent account of the facts. Therefore, adopting this assumption as a working hypothesis, we must now consider more minutely the exact law of force.

The law was elaborately investigated by Coulomb, in 1784, by means of his torsion balance. In this instrument a vertical wire attached to the torsion-head *h* (Fig. 147) carries a horizontal insulating arm, at the end of which a small metal disc *d* is fastened.

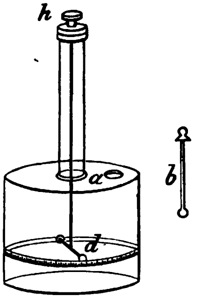


FIG. 147.

A scale fastened to the glass cover which surrounds the instrument enables us to determine the angular position of the arm; and the position of the torsion-head, when there is no torsion on the wire, is also noted. Now let a positive charge be given to the disc *d*, and let another positive charge, contained in a metal ball fixed to an insulating handle *b*, be introduced into the interior through the aperture *a* in the glass cover of the instrument, the length of the handle being such that the centres of the ball and the disc are in a horizontal plane. The mutual repulsion of the two quantities of electricity will cause the arm to twist round through a certain angle. Additional torsion is then put on the wire, by turning the head round, until the disc is brought back to its former position.

Now increase the charge in the ball in any ratio and repeat the same series of operations, having previously turned back the torsion-head into its old position. (A method of so increasing the charge will be discussed in § 284.) It will be found that the torsion, which must now be put on the wire in order to turn the disc back to its first position, is increased in the same ratio. This proves that the force is proportional to the quantity of electricity in the ball; and therefore, also, it is proportional to the quantity in the disc.

Next, perform a series of experiments in which the charges of both bodies are kept constant, while their mutual distance is varied. The amount of torsion which is requisite at the different distances will show that the force varies inversely as the square of the distance.

The same law will be found to hold when the charges are negative, and also when one is negative and the other positive. Of course, in the latter case, the angular rotations of the arm and the torsion-head are necessarily reversed in direction.

Let q, q', be the quantities, and let s be the distance between them. The law of force is expressed by

$$+\frac{qq'}{s^2},$$

where the positive sign corresponds to repulsion and so indicates that the force is attractive (*i.e.*, is in the direction of *decreasing* distance) when q and q' are of opposite sign, and that it is repulsive (*i.e.*, is positive outwards) when they are of like sign.

This is the well-known law of gravitational force, and therefore all results which can be deduced (Chap. VI.) regarding that force will at once apply to the case of electric force, provided that we take account of a possible reversal of sign. And conversely, theorems regarding the action of distributions of electricity have gravitational applications. Further, if it can be proved experimentally that there is no electric force in the interior of a charged sphere, the truth of the law of the inverse square follows at once. In this way Cavendish proved that the inverse power cannot differ from 2 by so much as one per cent. Maxwell, using modern instruments, proved that it did not differ from 2 by so much as one part in more than 40,000.

281. *Electric Potential. Electromotive Force.*—If we attempt to increase the charge of an insulated conductor by any stated means in which the same conditions are maintained, as in the electrophorus, we find that it is more and more difficult to do so the farther the process is carried out, and that the charge cannot be increased beyond a certain limit. To make the reason for this clear we must make an apparent digression.

Electrified systems obviously possess energy in virtue of their electrification, for the mutual attraction or repulsion between their various parts may be used for the production of mechanical work.

The *Mutual Potential Energy* of two systems is defined to be *the amount of work which may be obtained from their mutual repulsion until they are at an infinite distance apart;* and the *Potential* at any point, due to a given electrical system, is defined to be *the mutual potential energy between the system and unit quantity of positive electricity placed at that point.* This definition makes the sign of the potential coincide with the sign of the electrification of the system to which it is due.

Let V and V′ be the potentials at two points which are at a distance s apart. Since (§ 43) force measures the rate at which energy changes per unit of distance, and since it always acts so as to diminish potential energy, the average force which acts so as to transfer the unit of positive electricity from the point which is at potential V′ to the point which is at potential V is $(V'-V)/s$. When s is made very small, the average force becomes equal to the actual force at either point. The quantity $(V'-V)/s$ is then the rate of variation of potential per unit length, and is called the *Electromotive Force* at the given point, for it is the force which acts so as to transfer electricity.

We see therefore that no transference of electricity can occur between two conductors which are at the same potential.

Now we have seen that two like quantities of electricity, q and q', situated at a distance r apart, repel each other with a force

$$\frac{q'q}{r^2},$$

and so the force with which q acts on a unit of electricity of like sign is q/r^2. In order, therefore, to find an expression for the potential in terms of q and r, we have only to enquire what must be the values of V′ and V in order that we may have

$$\frac{V'-V}{r-r'}=\frac{q}{r^2},$$

when r and r' are practically equal. It is easy to see that we may take $V'=q/r'$, and

$$V=\frac{q}{r},$$

so that the value of the potential, due to a charge q, at a point whose distance from q is r, is the ratio of q to r.

The potential of a conductor must be uniform throughout if the electricity which it contains is at rest, for otherwise an electromotive force would act from one part of it to another. Hence we are justified in speaking of *the* potential of a conductor; and the above expression shows that the potential of a conductor becomes larger and larger in strict proportion to the quantity of electricity which it contains, being positive when the charge is positive, and negative when the charge is negative. And it follows that the work, which must be expended in order to bring up to a conductor a given charge, becomes greater and greater as the charge already contained in the conductor increases.

In particular, if the potential of the body (the disc of the electrophorus) which carries up the new charge does not exceed a certain fixed value, the potential of the conductor to which the charge is given cannot be made, by this process, to exceed that fixed limit. Therefore the *charge* of that conductor cannot be increased above a fixed limit, which is the statement made at the commencement of this section.

The potential is zero only at an infinite distance; but, in order to practically carry a charge to an infinite distance, it is necessary merely to connect the conductor, which contains it, to the ground. For the earth is a conductor which is practically at an infinite distance, and any ordinary charge which is communicated to it produces no sensible variation of its potential. (See next section.)

When the force is attractive, the potential—defined as above—becomes

$$V = -\frac{q}{r}.$$

The potential is therefore essentially negative; that is to say, work is done *against* the forces when the distance between the particles is increased. But it must carefully be observed that the fact of the potential energy being negative, at distances less than infinite, is due entirely to our having, for convenience, by making $V = q/r$, chosen infinite distance apart as the configuration of zero potential. The increase of potential energy is positive as we pass towards infinity.

The same thing holds when the force between two positive quantities is an attraction as it is in the case of gravitation. Still, to avoid the inconvenience of defining the gravitational potential as a negative quantity at all finite distances, it is preferable to change the sign and write, in this case also,

$$V = \frac{q}{r},$$

and

$$f = \frac{V - V'}{r - r'},$$

where f, being negative, signifies *inward* or *attractive* force. The potential V, therefore, does not represent the mutual potential energy. It is the *exhaustion* of potential energy which occurs when the unit of matter passes from infinity, under the attractive forces, to the position r.

The potential at any point, due to a number of separate masses, is simply the sum of the separate potentials due to each mass.

282. *Capacity. Condensers.*—The quantity of electricity which must be given to a conductor in order that its potential may be raised by unity is called the *Capacity* of that conductor.

For, since by symmetry the capacities of all similar portions of a spherical conductor must be equal, and since the potential must be constant throughout a conductor, the charge of a spherical conductor free from external electric influence must be symmetrically distributed with respect to the centre. Therefore, as the laws of electric and gravitational attraction are identical, we see that Newton's two theorems, stated in § 72, must hold in the case of a charged sphere. Thus, externally to the sphere, the action is the same as it would be if the charge were condensed at the centre, and so the potential of a sphere, of radius a, is

$$V = \frac{q}{a}.$$

Hence, when $V = 1$, $q = a$; that is, the capacity is measured by the radius. For this reason the capacity of the earth is large.

The capacity of a compound conductor which consists of two concentric spherical conductors may be made extremely great.

Let A and A′ (Fig. 148) represent the two surfaces, and let their radii be respectively a and $a + \tau$. Let A be charged with a quantity q of positive electricity, and let A′ be charged with an equal quantity of negative electricity. By reason of symmetry, as in the above case, these two charges will spread uniformly over the respective surfaces.

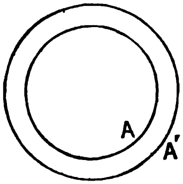


FIG. 148.

Now the potential at any point of A′, due to its own charge, is $-q/(a+\tau)$; and its potential, due to the charge of A, is $q/(a+\tau)$.

Its total potential is therefore zero; and so no flow of electricity will occur if it be connected to the ground by a conductor. This proves that if the sphere A be charged in any way, an equal and opposite charge will be induced on A′ if it be connected to the ground. Hence, if we add a positive unit of electricity to A, a negative unit will necessarily appear on A′ when it is 'put to earth'—to use the technical expression. But the work done in increasing the charge q of A by unity is q/a, and the work done in increasing the negative charge $-q$ of A′ by a negative unit is $-q/(a+\tau)$. Hence the whole work done is

$$w=\frac{q}{a}-\frac{q}{a+\tau}=q\frac{\tau}{a(a+\tau)}.$$

When τ is very small, this gives $q=a^2w/\tau$; and, when w is unity—that is, when the potential is unity—q represents the capacity of the arrangement, which is, therefore,

$$\frac{a^2}{\tau},$$

a quantity which may be made extremely great by sufficiently decreasing τ.

Any arrangement of this sort is called a *Condenser*, since it enables us, without much expenditure of work, to store up a large quantity of electricity. The only essential feature in any such arrangement is that the two conducting, and mutually insulated, surfaces shall be extremely close together in comparison with their own dimensions.

A common form of the condenser is known as the Leyden jar, since it was first employed by Muschenbrock, of Leyden. This consists (Fig. 149) of a thin glass jar coated externally and internally with tin-foil. The neck and upper portion of the jar are not coated with the foil, so that the insulation between the two conducting-sheets may be as complete as possible. The mouth of the jar is

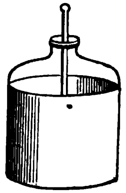


FIG. 149.

usually closed with a cork, through which passes a metallic rod, terminating externally in a knob, and making communication internally with the inner coating of the jar. The external coating can be readily connected to the ground, while the internal coating is charged, through the agency of the rod, by means of any electric machine.

Discharge of the jar is effected by means of the *discharging-rod* (Fig. 150), which consists of two jointed metal rods connected to a glass handle. The two knobs *a* and *b* are placed in connection with the outer and inner coatings of the jar; and the discharge, resulting in the combination of the two equal and opposite quantities of electricity, takes place through the metallic circuit *acb*. Great care must be taken in the use of jars charged to a high potential, as very serious, if not fatal, effects might ensue upon their discharge through the human body.

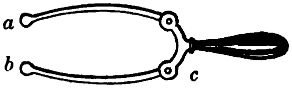


FIG. 150.

Another common form of condenser consists of a pile of sheets of tin-foil separated by paper soaked in paraffin. All the odd sheets in the pile (counting from one end) are connected together. So also are the even sheets, but the odd and the even elements are carefully kept separate. This arrangement constitutes a condenser of great capacity.

A practically identical arrangement may be made by joining together all the internal coatings and, independently, all the external coatings of a number of Leyden jars. The capacity of the whole is equal to the sum of the capacities of the separate jars.

If the jars be connected together in series—that is, with the outer coating of one joined to the inner coating of the next in order, and so on—the resultant capacity is only equal to the capacity of each jar (all being supposed to be equal in this respect), and the whole charge is equal to the charge of one jar when its coatings are raised to the same (total) difference of potential. When the capacities are not alike, we may let V_0 and V_1 indicate the potentials of the outside and inside coatings of the first jar, whose capacity is C_1, while V_1 and V_2 are the similar quantities for the second jar, the capacity of which is C_2, and so on, V_n being the potential of the inner coating of the last jar. If C be the total capacity, $C(V_n - V_0)$ is the total charge. But this is equal to the sum of the separate charges. Hence

$$C(V_n - V_0) = C_1(V_1 - V_0) + C_2(V_2 - V_1) + \dots\dots\dots + C_n(V_n - V_{n-1}).$$

But the charge of any one jar is necessarily equal to that of any other; for the outside coating of each has a charge which is equal and opposite to that of the inner coating of the jar to which it is joined, since the two form a single insulated conductor; and the charges in the two coatings of any one jar are also necessarily equal. Hence we have the $n-1$ equations

$$C_1(V_1 - V_0) = C_2(V_2 - V_1) = \dots\dots\dots = C_n(V_n - V_{n-1}).$$

In all there are n equations connecting the $2n+2$ quantities $C, C_1, \dots, C_n, V_0, \dots, V_n$. If V_0, V_n, and the quantities $C_1 \dots C_n$, are known, we can eliminate the Vs and calculate C.

In order to determine capacities readily, Cavendish constructed a series of condensers of different capacities. He determined the values of these capacities by comparison with the capacity of a sphere of known radius. If one of these condensers were joined with a condenser whose capacity was unknown, two of the coatings being joined together while the other two were connected to the ground, a charge (positive, say) given to the coatings which were insulated from the ground would be divided between the condensers in the ratio of their capacities. If the condensers were then separated and rejoined, so that the positively charged coating of one was connected to the negatively charged coating of the other, the charge of each would be destroyed if the capacities were equal. If the originally positively charged coating of the standard condenser was seen to be still positively charged when it was tested by means of a gold-leaf electroscope, the capacity of the standard condenser was greater than the unknown capacity; if it were negatively charged, the opposite conclusion held. In this way a pair of standard capacities, between which the unknown capacity lay, was determined.

Various modern methods of determining capacities are in use. In one, the coatings of the condenser are charged to a measured difference of potential, and the quantity of electricity required is found by discharging the condenser through a ballistic galvanometer (Chap. XXIX.).

283. *Specific Inductive Capacity.*—In last section we obtained the expression a^2/τ for the capacity of an arrangement consisting of two concentric conducting spherical surfaces of mean radius a and

separated by an insulating interval of thickness τ. If s be the surface of a sphere of radius a, the capacity is $s/4\pi\tau$. This may be taken as a close approximation to the value of the capacity of any condenser, such as a Leyden jar, if s be the area of one of the insulated surfaces.

It is customary to consider air under ordinary conditions as the standard insulating material.

If the air be replaced by some other insulating material, such as glass, resin, etc., it is found that the capacity of the spherical arrangement becomes

$$C = K\frac{a^2}{\tau}$$

where K is a constant for that particular medium and is called the *Specific Inductive Capacity* of the medium. The fact that the capacity is dependent on the nature of the insulating medium was discovered by Cavendish.

The Specific Inductive Capacity of a substance may therefore be determined by means of measurements of the quantity of electricity which is required to charge a jar, of which the given substance forms the insulating medium, up to a given potential. The ratio of this quantity to the quantity which is required in order to raise the potential of a precisely similar jar to the same extent, when air is the insulator, is the value of the required constant.

Faraday determined the Specific Inductive Capacity of various substances by measurements of potential. He used two precisely similar and equal condensers, each consisting of two concentric spheres which were so constructed that the insulating medium could be changed when desired. The coatings of one of the two condensers were insulated by air; those of the other were insulated by a substance whose (unknown) inductive capacity K, was to be found. He charged the air condenser to potential V and then divided its charge between itself and the other condenser by making connection between their outer coatings and their inner coatings respectively. V′ being the resultant common potential of the condensers, while C is the electrostatic capacity of the air condenser, the equation

$$V'C + V'KC = VC$$

expresses the condition that the total quantity of electricity in the two jars is equal to that originally possessed by the air condenser. Hence

$$K = \frac{V - V'}{V'}$$

is found in terms of the known potentials.

The value of this constant (sometimes called, after Maxwell, the Dielectric Constant) might also be found by observations of the reduction of the potential of a condensing arrangement, with a given charge, when the layer of air between the oppositely electrified surfaces is partially displaced by a layer of another insulating material. If t be the thickness of the new layer, the effective thickness of air displaced is t/K.

If vacuum be taken as the standard insulating medium, the specific inductive capacity of air is 1·00059, according to the determinations of Boltzmann. Faraday, with the apparatus at his disposal, was unable to observe any difference between the inductive capacities of different gases; but Boltzmann has shown that small differences really occur.

The dielectric constants of solids and liquids are greater than those of gases and differ considerably among themselves, as the following table shows:

Dense flint-glass ...	10·10	Turpentine	2·15
Beeswax	3·67	Bisulphide of carbon ...	1·81
Sulphur	2·58	Carbonic acid gas ...	1·000356
Paraffin	1·96	Hydrogen	0·999674

284. *Distribution of Electricity on Conductors. Electric Density.*—A static charge of electricity, which is communicated to a conductor, is necessarily confined to the surface of that body. For, otherwise, the mutual repulsion between like quantities of electricity would produce continuous currents of electricity in the interior of the conductor. Faraday showed that a conical gauze bag, whose rim was attached to an insulated ring, when strongly electrified on its outside, gave no trace of internal electrification. And on then turning the bag inside out by means of a silk cord, he showed that the same result held, thus proving that the electrification was always on the exterior. (Of course, as the bag is not a closed conductor, some electrification might be detected on the inner side near the rim.) Hence, if a charged conductor be placed inside a hollow insulated conductor which is then closed, and if the charged body be then put in contact with the hollow conductor, the charge will entirely pass to the outer surface, leaving the inner body uncharged. The body may then be withdrawn and the process be repeated.

The fact that the electricity is at rest on the surface of the conductor also shows that the distribution must be such as to produce a uniform potential all over the conductor.

If the conductor be spherical, we see, from the principle of symmetry, that the surface distribution of electricity is uniform;

but if the conductor be not symmetrical, we might infer that the quantity of electricity distributed, per unit of area, over the surface cannot be uniform. This quantity is called the *density* of the surface distribution. It is large where the curvature of the surface is large, and is small where the curvature is small.

In order to determine the density at various parts of any surface (whether conducting or not), we might place a small plane metallic disc, attached to an insulating handle, in contact with the surface. If the curvature of the surface be not too great, and if the disc be sufficiently thin and small, the electricity on the part of the surface covered by the disc will be transferred to the outside surface of the disc, for it is then practically a portion of the electrified body. The disc may then be removed, and the magnitude of its charge may be tested. If the charge be q, while the area is a, the density at that part of the surface is q/a. The process may be repeated as often as we please, at different portions of the surface, so long as the total charge of the conductor is not sensibly diminished; and this objection might be entirely avoided by taking precautions to have the potential of the conductor maintained constant. As this method was first used by Coulomb, the disc is called Coulomb's Proof Plane.

In a number of cases the law of distribution of density can be calculated from the known electrostatic laws. Some of these cases will be discussed in the immediately following sections.

285. *Electric Images.*—We have already seen (§ 282) that a uniform distribution of electricity over a spherical surface produces the same effect at external points as an equal quantity of electricity condensed at its centre would produce. This imaginary quantity is called the *electric image* of the uniform spherical distribution. [There is a reason for calling this quantity imaginary beyond the mere fact that it does not actually exist, and this is that the volume density of a finite quantity of electricity condensed at a point would necessarily be infinite, and such a condition cannot exist (§ 294).] We proceed now to the general discussion of the method of electric images, which is due to Lord Kelvin, and which gives, in many cases, simple solutions of very formidable problems.

Draw a sphere (Fig. 151) with radius CM $=a$, and divide CM externally and internally in the points A, A′ respectively, so that CA.CA′ $=a^2$. Take any point P in the circumference of the sphere, and join PA, PA′, PC. Let PA $=r$, PA′ $=r'$, CA $=d$. We know by geometry that the triangles CA′P, CPA are similar.

Therefore $$\frac{\text{CA}'}{r'}=\frac{a.}{r}$$

Now put $$\frac{e'}{r'} = -\frac{e}{r},$$

and we get $$e' = -e\frac{\mathrm{CA}'}{a}.$$

But if we draw A′Q perpendicular to CA, we have QA perpendicular to CQ, and therefore $\mathrm{CA}'/a = a/d$. Hence

$$e' = -e\frac{a}{d}.$$

If a quantity e of electricity be placed at A, while a quantity e' is placed at A′, the condition $e'/r' + e/r = 0$ asserts that the potential at the point P, due to these charges, is zero; and therefore the sphere is a surface of zero potential.

Consequently, if the sphere be a thin conductor, and be connected to the ground, the action of the charge e placed at A will induce upon it a charge, the effect of which at external points will be the same as that of e' placed at A′; and the charge e' placed at A′ will induce on the sphere a charge, the effect of which at internal points is the same as that of e placed at A. And, further, these two induced charges are precisely equal and similarly distributed, but of opposite sign; for, when the sphere is uncharged and insulated, and the charges e and e' are placed at A and A′ respectively, the potential of the sphere is zero, and no resultant charge is induced upon it, neither can electricity flow from any one part of it to any other part.

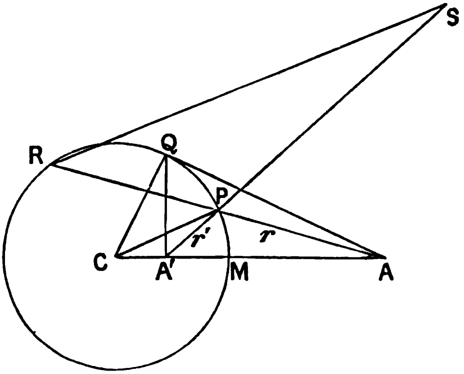


FIG. 151.

[To find the law of distribution of density which produces this effect, produce AP to R, describe a circle round the points R, A′, and A, and let A′P meet this circle in S. The triangles RPS and A′PA are similar. Also PR and PS are proportional to the forces, f and f', at P due to e and e' respectively. For these forces are e/r^2 and e'/r'^2 respectively, and $e/r = -e'/r'$. Hence the forces are inversely proportional to r and r', and therefore are directly proportional to PR and PS by construction. Consequently, the resultant force F at P is proportional to SR, and so $SR/SP = F/f' = AA'/r$.

This gives $$F = \frac{AA'}{r}f' = \frac{AA'e^2}{e'}\frac{1}{r^3} = \frac{AA'e'^2}{e}\frac{1}{r'^3}.$$

Hence a distribution of electricity on a spherical shell, having a density which is inversely proportional to the cube of the distance from an external (or internal) point, attracts internal (or external) matter, as if it were condensed at that point.]

The point A′ is the *image* of A with respect to the given sphere, and by means of the relation $CA.CA' = a^2$, we can find the image of any distribution of electricity.

For example, let there be a given distribution of electricity along the straight line A_1A_2 (Fig. 152). The image of the point A_1 in a sphere, of radius a, the centre of which is at C, is A'_1, which is such that $CA_1.CA'_1 = a^2$. Similarly, $CA_2.CA'_2 = a^2$, and so on. It is easy to prove that the line $A'_1A'_2$ is a portion of circle. But if A_1A_2 is very small $A'_1A'_2$ is practically straight, and is inclined to A_1C at an angle equal to the angle of inclination between CA_1 and A_1A_2.

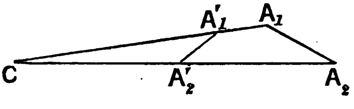


FIG. 152.

Let $CA_1 = r$, $CA'_1 = r'$, $A_1A_2 = l$, $A'_1A'_2 = l'$, and we get, as the ratio of an infinitely small length in the image to an infinitely small length in the direct system,

$$\frac{l'}{l} = \frac{r'}{r} = \frac{a^2}{r^2} = \frac{r'^2}{a^2}.$$

Similarly, α' and α representing areas, v' and v representing volumes,

$$\frac{\alpha'}{\alpha} = \frac{a^4}{r^4} = \frac{r'^4}{a^4};\ \frac{v'}{v} = \frac{a^6}{r^6} = \frac{r'^6}{a^6}.$$

Next let λ, σ, ρ, represent line, surface, and volume densities in the direct system, and let λ', σ', ρ' be the corresponding densities in the image. We get, considering magnitude alone,

$$\frac{\lambda'}{\lambda}=\frac{\frac{e'}{l'}}{\frac{e}{l}}=\frac{e'}{e}\frac{l}{l'}=\frac{a}{r}\frac{r}{r'}=\frac{a}{r'}=\frac{r}{a}.$$

Similarly, $$\frac{\sigma'}{\sigma}=\frac{e'}{e}\frac{\alpha}{\alpha'}=\frac{a}{r}\frac{r^4}{a^4}=\frac{r^3}{a^3}=\frac{a^3}{r'^3};\ \frac{\rho'}{\rho}=\frac{r^5}{a^5}=\frac{a^5}{r'^5}.$$

Again, if V and V′ represent direct and image potentials, we get

$$\frac{V'}{V}=\frac{\frac{e'}{r'}}{\frac{e}{r}}=\frac{e'}{e}\frac{r}{r'}=\frac{a}{r}\frac{r}{r'}=\frac{a}{r'}=\frac{r}{a}.$$

It is upon this relation, in terms of which *we can find the distribution of potential in the inverted system when we know the distribution of potential in the direct system*, that the physical use of th method of electrical images essentially depends.

For example, let a sphere be charged uniformly with electricity, and let us invert it with respect to itself. In that case, while the image coincides with the direct system, each point outside the image corresponds to a point inside the original sphere. But

$$V'=V\frac{a}{r'},$$

and we know that V is constant. Hence the potential at a point outside a uniformly charged sphere is inversely proportional to the distance of that point from its centre.

Next, invert the uniformly charged sphere with respect to an external point. The inverted system is also a sphere, and the density of the distribution upon it is given by

$$\sigma'=\sigma\frac{a^3}{r'^3},$$

where σ is constant. Hence the density of the image sphere varies inversely as the cube of the distance from the point of inversion. Further, points inside the first sphere invert into points inside the image. Hence the potential equation shows that the potential at an internal point varies inversely as the distance of that point from the centre of inversion, so that internal electrification is attracted as if the spherical distribution were condensed at the centre of inversion.

Finally, invert the uniformly charged sphere with respect to an internal point. The density of the image sphere varies inversely as the cube of the distance from the centre of inversion, and the potential varies inversely as that distance, but internal points have inverted into external points. Hence electricity at an external point is attracted by a spherical shell, whose density varies inversely as the cube of the distance from an internal point, as if the spherical distribution were condensed at that point.

These three propositions will be otherwise proved in the immediately succeeding sections. A comparison of the present proofs with those there given will exhibit to a slight extent the great power and simplicity of the method of electric images.

286. *Equipotential Surfaces. Lines and Tubes of Force.*—Any surface over which V is constant is called an Equipotential Surface; and a line, the tangent to which at any point is always in the direction of the force at that point, is called a *Line of Force.* Each line of force is a possible path along which electricity would move under the given forces, and its positive direction is chosen as that in which positive electricity would be urged.

Since the change of potential is zero as we pass from one point to another point of an equipotential surface, it follows that the force at any point of such a surface has no component along it. In other words, the lines of force are everywhere perpendicular to the equipotential surfaces. They proceed from a point at high potential to a point at low potential, originate at a positive charge, and end at a negative charge, or pass to infinity.

No two different equipotential surfaces can intersect one another; for this would imply that a finite difference of potential exists between points at an infinitely small distance apart, *i.e.*, the force would be infinite at any such intersection, and no examples of infinite forces occur in nature.

If, through every point of an infinitely small closed curve drawn on an equipotential surface, we draw lines of force, a tube of infinitely small section will be formed. Such a tube is called a *Tube of Force.*

Necessarily, the lines and tubes of force can originate only at places where electricity is situated; for the force owes its origin to the presence of electricity. And if, at any point of space, a force f exists, we may draw an infinitesimal tube of force, so as to contain that point and to enclose f lines of force per unit area of its normal section. The number of the lines of force per unit of sectional area then indicate the intensity of the force at the given point.

If a given portion of a tube does not contain any electricity, the number of lines which it contains remains constant, since no line can end in its interior, and none can pass out through its sides. This gives

$$f\sigma = c,$$

where f is the force at any point of the tube at which the sectional area is σ, and c is a constant, numerically equal to the number of lines which the tube contains.

287. *Special Applications.*—(1) Let the distribution of electricity be symmetrical about a point. In this case the tubes of force are cones, and the area of any normal section of each of them is proportional to the square of its distance from the point of symmetry. Hence the above equation shows that the force at any point is inversely proportional to the square of the distance of that point from the point of symmetry, say

$$f = \frac{a}{s^2}.$$

Of course this result merely justifies, for the law of the inverse square, the above use of the lines and tubes of force. If, for example, the force varied in inverse proportion to the fourth power of the distance, we would have to use the square of the number of lines of force per unit area to represent the actual value of the force.

(2) Let the electricity be symmetrically arranged about an infinitely long straight axis. The equipotential surfaces are concentric cylinders whose common axis coincides with the axis of symmetry, and the tubes of force are wedges bounded by axial planes. The sectional area is proportional to the distance from the axis, and therefore the force is inversely proportional to that distance, *i.e.*,

$$f = \frac{b}{s}.$$

(3) Let the electricity be arranged homogeneously in infinite parallel planes. The equipotential surfaces are planes parallel to these, and the tubes of force are cylinders arranged perpendicularly to the planes. Therefore the force is constant at all distances, say

$$f = c.$$

288. *Total Force over a Closed Surface.*—Draw any closed surface S (Fig. 153), and,

(1) Let a quantity q of electricity be placed at a point m outside. Draw any infinitesimal tube of force, mnp, cutting the surface at n and p. (Of course, it may cut the surface in any *even* number of places.) The total force over the portion of the surface intercepted by the tube at n, is equal to that over the portion intercepted at p; but, in the one case, the force is directed outwards over the surface,

while, in the other, it is directed inwards. Hence, by consideration of an infinite number of such tubes of force intersecting all parts of the closed surface, we see that the *total outward force over the whole closed surface, due to a material particle situated at an external point, is zero.* [The use of the word *outward*, of course, implies that the force is to be reckoned as negative when it is inwardly directed.]

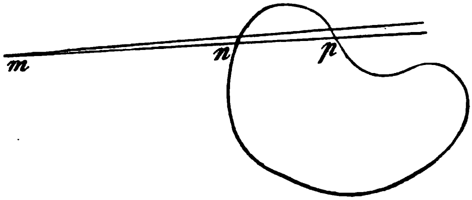


FIG. 153.

The same proof applies in the case of any number of electrified points.

(2) Let the point m, at which the quantity q of electricity is placed, be within the closed surface. Draw a sphere, with unit radius, from the point m as centre. The area of this spherical surface is 4π, and the force at any point of it, due to the repulsion of the central charge, is equal to q. The total outward force is therefore $4\pi q$. But, by the result of § 286, the total outward force over this surface is equal to that over the given surface S. Hence, if q be the whole amount of electricity contained within S, *the total outward force over the whole closed surface, due to electricity of amount* q *enclosed within it,* is $4\pi q$.

289. *Special Applications.*—We may apply these results to the determination of the values of the constants a, b, and c, in § 287.

In example (1) of that section, the whole number of lines of force which cross any closed surface surrounding the spherical distribution of electricity is $4\pi q$, where q is the whole amount of electricity. Hence, if we suppose the enclosing surface to be a concentric sphere of radius s, we get as the total force

$$4\pi s^2 f = 4\pi s^2 \frac{a}{s^2} = 4\pi q.$$

Therefore

$$a = q.$$

Similarly, if, in example (2), we consider the force due to the charge q contained in unit of length of the cylindrical distribution, we get

$$2\pi s f = 2\pi s \frac{b}{s} = 4\pi q,$$

for $2\pi s$ is the area of unit of length of a concentric cylinder of radius s.

Therefore $$b = 2q.$$

Again, if—example (8)—we draw a right circular cylinder of unit radius, perpendicular to the infinite planes, and close its ends by parallel planes on opposite sides of the given plane distribution of electricity, the force exerted over the two ends is

$$2\pi f = 2\pi c = 4\pi q,$$

so that $$c = 2q.$$

In particular, if the given distribution consists of an infinitely thin plane layer, of infinite extent, and of finite surface density σ, the force at any point is

$$f = 2q = 2\pi\sigma.$$

[Though such a distribution cannot occur in the case of gravitational matter, we can apply the result to the case of a plane layer of ordinary matter, of finite thickness θ and volume-density ρ, by writing $\sigma = \theta\rho$.] It follows that the force at any point *just outside* a surface on which electricity is distributed is normal to the surface and equal to $2\pi\sigma$—provided that the surface is of finite curvature; for the point may be taken so close to the surface that, as seen from it, the surface is practically an infinite plane; that is to say, any infinitely small portion of the surface is plane, and the point may be taken infinitely close to this portion in comparison with its dimensions, infinitely small though they be.

290. In the case just considered, the normal forces at the two sides of the surface are respectively $f = 2\pi\sigma$, and $f' = -2\pi\sigma$. The total change of force in crossing the surface is therefore

$$f - f' = 4\pi\sigma.$$

This shows us how *to distribute electricity over a given surface in order to produce a given change in the value of the normal force in passing from one side to the other.*

It is easy also to obtain an expression for the volume-density of electricity which is required to produce a given continuous distribution of force.

Take three rectangular axes at any point o (Fig. 154), and draw a little parallelepiped at this point with its edges parallel to the x, y, and z axes; and let the lengths of the edges be very small and equal to ξ, η, ζ, respectively.

Resolve the forces in the neighbourhood of the parallelepiped

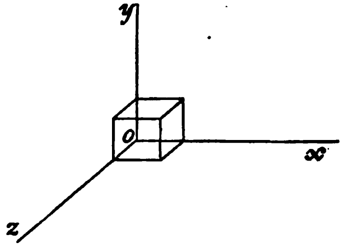


FIG. 154.

into their components parallel to the axes. This will not alter the result of § 288; and we may now draw the lines of force perpendicular to the various small surfaces.

Let n_x be the number of lines of force which cross unit area of the face of the parallelepiped which passes through the origin and is perpendicular to the x axis. The total number of lines which cross that face is therefore $n_x\eta\zeta$, since $\eta\zeta$ is the area of the face. Similarly the number which cross the parallel face at the distance ξ from the former is

$$(n_x+n'_x\xi)\eta\zeta,$$

where n'_x is the rate at which n_x varies per unit of length measured along the x axis. It must be remembered that n_x really represents the *excess* of the number of lines which cross in the direction ox over those which cross in the reverse direction. Hence the total number of lines which leave the little volume from within by the two faces is

$$n'_x\xi\eta\zeta,$$

the difference of these two quantities. By similar reasoning for the other pairs of faces we find that the total number of lines which leave the little volume (that is, the excess of those which leave over those which enter) is

$$(n'_x+n'_y+n'_z)\xi\eta\zeta,$$

But by § 288 this is equal to $4\pi\rho\xi\eta\zeta$, where ρ is the density of the matter contained in the volume $\xi\eta\zeta$. Therefore

$$(n'_x+n'_y+n'_z)=4\pi\rho.$$

And the result is independent of the size of the little volume; so that the meaning of the equation is that the volume density at any point of space is $1/4\pi$ times the number of outwardly drawn lines of force which originate, per unit of volume, at that point.

291. It is known from experiment that any closed conductor, which has an insulated charge placed inside it, and is connected to the ground, will become charged in such a way that no force is exerted at any external point by the internal and superficial charges. The reason is that the potential of the conductor is zero, so that the potential is uniformly zero in all space external to it when there is no external electrification. The tubes of force which proceed from, or end at, the internal charge, end on, or proceed from, the inner surface of the conductor. And this is true, whatever be the distribution of electricity, positive or negative, in the interior. This constitutes a proof of the important proposition that *It is possible so to distribute electricity over a given surface, which encloses a given distribution of electricty, as to produce, outside that surface, the same potential as that which the given mass produces.* For, we have only to suppose the induced surface distribution to be fixed and its sign to be reversed, in order to get, when the internal charges are removed, the same external effect as these internal charges themselves caused.

In the same way we see that the closed conductor will screen internal points from the effects of external charges, and that it is possible to obtain a surface distribution of electricity on the conductor which will produce, at internal points, the same effects as those which are produced by the given external charges.

Since, in the former case, all the lines of force which originate at the parts of the system must end (except in so far as they may proceed from one part of the system to another) upon the surrounding conductor, the charge induced on the conductor is equal in amount and opposite in sign to the total algebraic sum of the various internal charges. A particular case of this was discussed in § 282.

It is extremely difficult, or even impossible, in most cases to calculate the distribution of electricity which will give rise to a given distribution of potential. But the method of electrical images enables us to deduce with great ease the solution of many unknown problems from the known solutions of others.

292. *Electric Induction. Tubes of Induction.*—Hitherto we have spoken of gravitational and electrical action as if it took place directly at a distance. We have spoken of the mutual potential energy of two systems without inquiring *how* one system possesses energy in virtue of its position relatively to another. But, if we

believe that energy is transferred by means of matter (§ 7), we must look upon some intervening medium as the vehicle through which it is transferred.

This is the way in which Faraday regarded the subject; and a great part of the development of electrical science in recent times is due to Faraday's work together with Clerk-Maxwell's mathematical interpretation and development of his views. The fact, stated in section 283, that the electrical condition of charged conductors depends upon the nature of the intervening insulating medium gives strong support to the belief.

In Chap. XXVI. we shall see that, when a current of electricity flows along a conductor, the amount of electricity which crosses any section of the conductor, per unit of time, is constant. In other words, the flow of electricity resembles the flow of an incompressible fluid. Similarly, the facts that no quantity of one kind of electricity can appear without the simultaneous development of an equal quantity of the opposite kind, and that the quantity of electricity induced on a closed conductor, which entirely surrounds the inducing electricity, is equal in quantity, and opposite in sign, to the latter, indicate that the induction of electricity through a dielectric resembles the displacement of an incompressible fluid. [This fact, probably, prolonged the use of the objectionable term 'electric fluid.']

We must therefore look upon a conductor as a body which cannot sustain electrostatic stress, and so permits electricity to flow along it when such stress is applied; and we are to regard an insulator as a body which can sustain electrostatic stress, in which 'displacement' of electricity takes place in proportion to the stress which is applied, and in which the displacement is annulled when the stress is removed. In this way of looking at the matter, a surface charge is supposed to reside on the surface of the dielectric and not on the surface of a conductor.

The word 'displacement' was introduced by Maxwell from the analogy to an elastic medium the parts of which suffer displacement when stress is applied and recover from the distortion when the stress is removed. But Maxwell was very careful to avoid attaching any exact meaning to the term 'electric displacement.' He merely used it by analogy; and this cannot be too carefully kept in view.

Faraday used the term *Electric Induction* to indicate the state of the medium in virtue of which equal and opposite quantities of electricity appear on opposed surfaces. He illustrated his idea of the polarized state of the medium by the action of small silk filaments placed in turpentine contained in a glass vessel through opposite sides of which projected inwards two pointed metal rods, one of which was connected to the ground while the other was electrified. The filaments arranged themselves longitudinally in the neighbourhood of the line joining the rods.

Maxwell spoke of the total amount of induction through a given surface as the amount of electricity which is 'displaced' through it. A forward displacement in an insulator corresponds to a direct current in a conductor; a diminution of displacement corresponds to a reverse current.

The difference between the specific inductive capacities of various substances is explained by a difference in the amount of displacement which is produced in each under the same electromotive force.

As we have already found, we may draw lines of force through all points of any small closed curve on a conductor so as to form a tube of force; and we may draw such tubes, covering the whole surface of the conductor, in such a way that the number emanating per unit of area from all parts of the surface is equal to $4\pi\sigma$ where σ is the density of the electrification. The number of such tubes, which intersect unit area of any equipotential surface, therefore expresses the intensity of the force at that part of the equipotential surface.

But, instead of proceeding in the above manner, we may draw the tubes so that each encloses unit amount of electrification on the conductor. Faraday called such tubes *Tubes of Induction;* for, where they originate, they enclose unit quantity of positive electricity, and, where they end, they enclose unit quantity of negative electricity. The total number of tubes of induction originating from, or ending on, a conductor, express its total positive, or negative, electrification, and the 'induction,' or 'displacement,' through each section of such a tube is constant.

Properly speaking, tubes of induction are formed by *lines of induction*, and not by lines of force. And it is well to remember that tubes of induction are not necessarily tubes of force: for the displacement does not always take place in the direction of the electromotive force, although in general it does. (Compare the elastic properties of non-isotropic solids, § 219.)

293. *Electric Energy.*—In order to estimate the amount of energy which is associated with the charge of a conductor at a given potential, we have merely to calculate the work expended in charging it. Let Q be the final charge, and let V be the potential. Let us suppose that the conductor is charged by successive equal in-

finitesimal instalments, and that the charge at any instant is q, while the capacity of the conductor is C. The potential at the given instant is therefore q/C. But the potential is the work which is required in order to bring up unit charge from an infinite distance, and give it to the conductor. Hence the amount of work which is necessary in order to increase the charge q by a given small quantity is the product of that quantity into the potential; and the total amount of work which must be expended in raising the charge from zero to its final value is the product of the whole charge into the average value of the potential. Now, the potential changes at a constant rate when the charge increases at a constant rate. Therefore the average value of the potential during the process of charging is one half of the final value of the potential. The work is consequently

$$\tfrac{1}{2}QV = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\frac{Q^2}{C}.$$

That is, the whole energy is one half of the product of the charge into the potential; or one half of the product of the capacity into the square of the potential; or one half of the quotient of the square of the charge by the capacity.

Now let us look at the problem from the point of view of induction. Consider a positively electrified body inside a closed conductor. Draw unit tubes of induction from the body to the internal surface of the conductor, and describe equipotential surfaces corresponding to all potentials which differ from each other by unity, and are included between the potential V of the electrified body and the potential V′ of the surrounding conductor. The number of cells into which the tubes of induction and the equipotential surfaces divide the volume between the two charged surfaces is equal to the product of $V - V'$ into the number of tubes, *i.e.*, into the charge of the body. Hence the number of cells into which the space is divided is double of the electrical energy of the system.

A simple extension of this reasoning shows that the same result is true whatever be the number of electrified bodies contained inside the conductor. (See Maxwell's *Elementary Treatise on Electricity*, Chap. V.)

The result might suggest that the energy of a system of charged conductors is contained, not in the conductors themselves, but in the insulating medium which surrounds them. And Faraday's and Maxwell's views of the nature of induction show us how this may be. The dielectric is in a state of strain so long as displacement is maintained by the action of electromotive force; so that the energy which was expended in producing the strain is contained in the dielectric in a potential form. Corroborative evidence is found in the fact that visible strain may be produced in a piece of glass by means of electrostatic stress.

To obtain an expression for the amount of energy contained, per unit of volume, in the dielectric, let us consider, as the simplest case, an insulated sphere of variable radius r charged with a constant quantity q of positive electricity. The potential of the sphere is q/r, and the energy contained in the space external to it is

$$\tfrac{1}{2}q\,.\frac{q}{r}.$$

Now let r increase infinitesimally to $r+\rho$. The energy becomes

$$\tfrac{1}{2}q\,.\,\frac{q}{r+\rho}.$$

The difference of these quantities,

$$\tfrac{1}{2}q^2\frac{\rho}{r^2},$$

is the energy which is contained in the intervening shell of volume $4\pi r^2\rho$. Hence the energy contained per unit of volume at the distance r is

$$E=\frac{1}{8\pi}\,\frac{q^2}{r^4}=\frac{1}{8\pi}F^2$$

where F is the resultant force at the distance r due to the charge q. If K be the specific inductive capacity of the medium we must write

$$E=\frac{K}{8\pi}F^2.$$

This result is quite general, F being the resultant force at any point due to the total electrification.

Maxwell has investigated the nature of the stress in the dielectric which would account for observed electric phenomena. He finds that the stress consists of a *tension* $KF^2/4\pi$ along the lines of force coupled with an equal *pressure* in all directions at right angles to the lines of force.

In particular, the tension at the charged surface of a conductor is $4\pi K\sigma^2$ in a direction perpendicular to the surface, where σ is the surface density of the electrification.

As an instructive example in connection with the above expression for the energy of an electrified system, we may estimate the energy of a charge Q, first when it is contained in a jar of capacity

C, and, second, when it is divided between that jar and another jar of capacity C′. The original energy is

$$\frac{1}{2}\frac{Q^2}{C}.$$

After division, since the potentials of the two jars are equal, we have

$$\frac{Q_1}{C}=\frac{Q_2}{C'},$$

where Q_1 and Q_2 are the charges in the jars of capacities C and C′ respectively. Also

$$Q_1+Q_2=Q,$$

and hence $$Q_1=Q\frac{C}{C+C'},\ Q_2=Q\frac{C'}{C+C'}.$$

And the respective energies are

$$\frac{1}{2}\frac{Q_1^2}{C}=\frac{1}{2}Q^2\frac{C}{(C+C')^2},\ \frac{1}{2}\frac{Q_2^2}{C'}=\frac{1}{2}Q^2\frac{C'}{(C+C')^2}.$$

Therefore, the total energy is

$$\frac{1}{2}\frac{Q^2}{C+C'},$$

which is always less than the original energy. The fraction $C'/(C+C')$ of the whole energy has been dissipated in the process of division—usually taking the form of sound, light, and heat. [Compare the dissipation of energy, which takes place when a gas is allowed to expand without doing work, as in Joule's experiment (Chap. XXXI.), or that which occurs when heat diffuses, so as to arrive at a lower temperature without the performance of work. In the present case no electricity is lost, but the potential is lowered.]

294. *Electric Absorption. Disruptive Discharge.*—If an elastic medium be distorted beyond its limits of perfect elasticity, the removal of the stress is not followed by complete recovery from strain; but, if the distortion be not too great, complete recovery may take place after a sufficient time has elapsed. Conversely, a long-continued force may produce large distortion.

Analogous phenomena appear in dielectric media when subjected to electrostatic stress.

Thus, a Leyden jar, when charged to a certain potential, will gradually fall in potential, though it is well insulated. The result is the same as if its capacity gradually increased, or as if the specific inductive capacity gradually increased so that the same displacement was maintained by a smaller difference of potential. If the jar be now discharged, the quantity of electricity which is obtained is smaller than the original charge. This phenomenon is known as *Electric Absorption*, for the jar *appears* to have absorbed some of its charge.

A second (small) discharge may be obtained from the jar if it be left for some time. Subsequently a third may be obtained, and so on until the total discharge equals the original charge. These are called *Residual Discharges*. The apparently absorbed charge seems to leak out again, as if the medium gradually recovered from a temporary distortion.

It is not well to pursue these analogies too far. Maxwell has shown that apparent absorption will take place if the insulating medium is heterogeneous in the sense that it consists of parts the specific inductive capacities of which differ from one another, or of parts which differ from each other in their insulating power. The insulating power of ordinary dielectrics, such as glass, gutta percha, etc., is not perfect; and so, if a composite dielectric consisted of alternate layers of incompletely and completely insulating materials, electric absorption would be manifested.

We know also that elastic solids are only capable of withstanding strain to a limited extent, and that they will be ruptured if too great stress be applied. Similarly all dielectrics will cease to insulate electricity if they are subjected to too great electrostatic stress. The state of strain in the insulating material of a Leyden jar becomes greater and greater as the potential of the jar is raised higher; and, if the process be continued too far, the insulation breaks down, and the separated electricities recombine through the ruptured dielectric. This phenomenon is called the *Disruptive Discharge*.

A Leyden jar, through the substance of which the disruptive discharge has occurred, is useless for all subsequent electrical purposes, for the glass is in part shattered by the discharge. On the other hand, if air or any other fluid were used as the dielectric, the jar would insulate as completely as ever it did so long as too great stress were not again applied; for, although by the energy of the discharge the parts of the fluid medium would be violently disrupted, the insulation would be restored by an inflow of the surrounding medium.

The disruptive discharge is usually accompanied by the production of sound, light, heat, and mechanical effect, the total energy evolved being the exact equivalent of the original electrical energy.

Various forms of the disruptive discharge exist. The most ordinary form is called the *spark* discharge. When two oppositely charged surfaces are brought sufficiently near each other, the electrostatic stress in the medium increases to such an extent that the electricities combine because of rupture of the dielectric between the charged surfaces. A small streak of light is apparent where the discharge occurs, its form depending upon the thickness of the dielectric through which the discharge occurs. When the distance is great the streak of light (the spark) is very irregular and jagged in outline.

Feddersen found that the nature of the spark discharge depends upon the resistance (Chap. XXVI.) of the circuit in which the discharge occurs. When the resistance is sufficient, the spark consists of successive rapid discharges in the same direction. It apparently becomes continuous when the resistance is lessened to a certain extent, and it consists of a rapidly alternating series of discharges in opposite directions, when the resistance is still further diminished. The existence of an oscillatory discharge had previously been inferred by Henry from the fact that needles which were magnetized by a disruptive discharge did not always exhibit the same polarity. Under ordinary conditions the spark lasts for a few hundred millionth parts of a second.

Sometimes the discharge is in the form of a *brush*. This is seen chiefly when one of the two conductors has great curvature at the place where the discharge occurs. A short line of light, which abruptly branches out into a brush-like form, appears at the place. Wheatstone showed, by means of his revolving mirror, that the brush discharge consists of a series of intermittent and separate discharges. Its intermittent character gives rise to the crackling, or even musical, sound which accompanies this form of the discharge. Faraday considered that the brush discharge takes place from the electrode to the air.

The *glow* discharge takes place from the rounded extremity of a wire which projects into the air. The end of the wire is covered by a phosphorescent light. This form of the discharge does not appear to be intermittent. It seems rather to be, as Faraday concluded, a convective discharge, in which the charge is carried away by the air. (Compare the action of the pith-ball, § 275.)

The 'electric wind,' which blows from a sharp electrified point, is due to the repulsion of air which has been electrified by contact with the point.

The limiting tension (§ 293) which the insulating medium can sustain without rupture is called the *Dielectric Strength* of the medium. The dielectric strength of air depends upon the distance between the oppositely electrified surfaces. It has a greater value when the distance is small than it has when the distance is large. In all gases it increases as the pressure increases, and diminishes as the pressure diminishes—but not indefinitely. A minimum value is reached at a certain stage, beyond which the strength increases as the pressure is farther diminished.

The method of spark discharge under diminished pressure (in so-called vacuum tubes) is much used for the purpose of examination of the spectra of gases. This subject will receive further treatment in Chap. XXVII.

295. *Atmospheric Electricity, etc.*—The atmosphere is almost always in a state of electrification, either positive or negative. The electrification is generally positive during long-continued fine weather; it generally becomes negative when the fine weather breaks.

In order to test the nature of the electrification, use may be made of Kelvin's water-dropping accumulator. This instrument consists of an insulated metallic vessel, which contains water, and which is fitted with a long fine nozzle from which the water issues drop by drop when the stopcock with which it is fitted is opened. The nozzle projects out into the external atmosphere by an opening in a window, and the vessel is connected with an electrometer. The stopcock is then opened and the water drops out.

In Kelvin's portable electrometer, designed specially for the investigation of the electrification of the atmosphere, the fumes from a burning match perform the function of the water drops in the above arrangement.

If the atmosphere be positively electrified, negative electricity will be induced in the nozzle and, therefore, in the drop, while positive electricity is repelled to the electrometer. As each drop falls away, carrying its negative charge with it, the vessel and electrometer are left more and more positively charged. The electrification of the atmosphere is, therefore, indicated by the development of a charge of like sign in the electrometer.

An interesting question arises in this connection—What is the source of the energy of the charge in the electrometer? The energy of the charge may be transformed into heat, and the energy of the falling drops may also be transformed into heat. Further, there is no other possible source of heat in the arrangement. But the drops may fall without any production of electric charge, and, therefore, the principle of conservation compels us to assert that the drops will fall more slowly when they are electrified than they do when unelectrified, and so will do less work. This conclusion is verified by experiment.

If we replace the metallic vessel (above alluded to) by a hot crucible into which we drop water, the water will evaporate, and the vapour will be found to be negatively electrified, while the crucible and the electrometer become positively charged. If the vapour condenses, the total volume of all the drops of water which are formed remains constant; but the total surface of the drops diminishes as each drop increases in size, and the same quantity of electricity is confined to a smaller surface. The result is that the potential of the drops rises considerably. It is possible that the high potential of thunder-clouds may be explained in this way.

When the potential rises to such an extent that the air is unable to withstand the electrostatic stress, disruptive discharge (lightning) takes place.

The great use of a lightning-rod is to prevent the potential from rising to such an extent that disruptive discharge will occur. It does this by drawing off from the surrounding air a continuous current of electricity. The electrified air induces the opposite electrification in the rod, and the density is very great at the sharp point—so great, that the electricity streams off from it to the air by silent discharge, and so annuls, totally or partially, the electrification of the air; and this is equivalent to the passage of the opposite electricity from the air to the ground through the rod. If a cloud in the neighbourhood of the rod were suddenly electrified to a high potential by disruptive discharge from a distant thunder-cloud, the rod might not be able to draw off the electricity with a sufficient rapidity to prevent discharge from the near cloud to the building supposed to be protected by the rod. The rod would, more likely than not, be insufficient for the purpose of carrying off the discharge to the ground.

296. *Pyro-electricity. Piezo-electricity.*—If crystals of tourmaline, or of some other minerals, be heated, electrical phenomena will be manifested, although previously the crystals appeared to be unelectrified. Positive electricity appears at one end of the crystallographic axis, negative electricity appears at the other.

The electrification may then be destroyed by passing the crystals through a flame. Further electrification will then be manifested, similar to that which formerly appeared, if the heating be proceeded with; but if, on the contrary, the crystals be allowed to cool, the opposite electrifications will appear at the ends.

Lord Kelvin supposes that such crystals possess internal electrification—that they are electrically polarised in the direction of the axis—and that, when they are passed through the flame, their surfaces become electrified in such a way as to annul at all external points the effect of the internal electrification. And he further supposes that the amount of internal electrification depends upon the temperature, so that heating or cooling disturbs the balance of external effects.

Let a plate be cut from a hexagonal prism of quartz. Let its length be perpendicular to any one pair of the six plane sides of the prism, while its breadth is parallel to the six longitudinal edges of the prism. If longitudinal tension be applied to this plate, the sides become oppositely electrified. The sign of the electrification is reversed if pressure be applied instead of tension. These facts may be readily shown by silvering the opposite faces of the quartz and connecting them to an electrometer.

Lord Kelvin points out that electric aeolotropy of the molecule, and nothing but electric aeolotropy of the molecule, can explain the phenomenon. He has shown that the results follow from the supposition that the crystalline molecule consists of some configuration of three positively charged atoms of silicon and three negatively charged double atoms of oxygen.

297. *Electrification by Contact.*—The electrification of glass or of sealing-wax, etc., by friction may be explained by the assumption that an electromotive force exists at the surface of contact of the two substances which tends to produce electric displacement across the interface, and that friction is used merely for the purpose of securing better contact.

The substances being non-conductors, the electricity cannot pass from the surface, and the displacement continues until the effect of the reverse force which it entails balances the effect of the electromotive force of contact. So long as the surfaces remain in contact, the arrangement acts as a condenser of extremely large capacity, and relatively large displacement may be produced by a comparatively feeble difference of potential. But whenever the surfaces are separated, the potential rises greatly, because of the ensuing decrease of capacity. In this way the high potential obtainable from frictional machines, or from the electrophorus, etc., is explained. The charge is the same after separation as before it, but the potential has increased, and therefore the energy has increased; and the increase of energy is the precise equivalent of the work done in the process of separation.

Now, although we cannot electrify conductors by friction in the same way as we electrify non-conductors, we can produce electrification of conductors by contact or friction, provided that we take proper precautions.

If we take two flat pieces of zinc and copper, insulate them both, and then place their flat faces in contact, the copper will become negatively electrified, while the zinc becomes positively electrified. This may be proved by separating the plates (still insulated) and testing their electrification by means of the electroscope or the electrometer. And we may explain the result by stating that an electromotive force of contact acts at the surface of separation of the metals in the direction from copper to zinc.

Volta found that this assumed electromotive force of contact between any pair of metals is equal to the sum of the electromotive forces between every pair of metals forming a series closed by the given pair. From this it would follow that the sum of the contact forces in any complete heterogeneous metallic circuit is zero. This is known to be true so long as the temperature is uniform throughout the circuit; and it is in accordance with the principle of conservation of energy, for there is no source of energy in such a circuit.

The assumption of the existence of an electromotive force of contact between metals sufficiently great to account for the observed effects is regarded as inadmissible by many physicists. That a contact force does exist is shown by thermo-electric phenomena; but this force is very much smaller than the Volta contact force. Consequently, those physicists who are unwilling to admit the possibility of a true contact force between metals, which would account for the whole observed effect, look upon the surfaces of separation between the metals and the air as the real seat of the electromotive force. The whole question is still involved in much uncertainty.

Contact forces exist between metals and liquids, and also between different liquids. Volta's law does not hold universally in the latter case.

298. *The Electrometer.*—Instruments such as the gold-leaf electroscope may be used for the purpose of obtaining very rough measurements of electromotive force. The *electrometer*, in one or other of its various forms, is used when accurate measurements of electrostatic effects are required. It is used directly for the determination of difference of potential and it may be indirectly used for the purpose of the comparison of the capacities of conductors, and consequently for the determination of their charges.

Instruments such as the gold-leaf electroscope are termed *idiostatic* instruments, for there is no electrification in any part of these instruments except such as is due to the electrification which is to be tested; and their indications (when small) are therefore proportional to the square of the difference of potential which is to be observed. Any small variation of potential is therefore inappreciable when the potential itself is small, and the indications of the instrument are of the same kind, whether the potential is positive or negative. In *heterostatic* instruments, some of which we shall now describe, a constant charge of one definite kind is maintained in one part of the apparatus, so that a small variation of potential produces the same effect, whether the potential is large or small, and the indication is reversed in direction when the potential changes sign.

Most forms of the electrometer depend for their action upon the electrostatic force between similarly or oppositely charged bodies.

Coulomb's torsion balance is therefore one (but a very imperfect) form of electrometer.

In the *attracted disc* electrometer the two charged bodies are in the form of parallel horizontal discs placed at a distance apart which is small in comparison with their transverse dimensions. We shall assume that the discs are oppositely charged, the densities of the electrifications being $+\sigma$ and $-\sigma$ respectively. Except in the near neighbourhood of the edges, the lines of force are perpendicular to the discs, and the force at any point between them is (§ 289) $2\pi\sigma - 2\pi(-\sigma) = 4\pi\sigma$ in the direction from the positively charged disc to the negatively charged disc. Since $2\pi\sigma$ is the force with which the positively charged disc acts upon unit quantity of negative electrification on the other disc, the total force with which it attracts that disc is $2\pi\sigma . \sigma a = 2\pi\sigma^2 a$, where a is the area of the disc. But, as above, the force at any point between the discs is $4\pi\sigma$, and is equal to $(V - V')/t$, where V and V′ are respectively the potentials of the positively and the negatively charged discs, and t is the interval between them. Hence $\sigma = (V - V')/4\pi t$, and the total force of attraction between the discs is

$$F = \frac{1}{8\pi} \frac{(V - V')^2}{t^2} a,$$

which gives
$$V - V' = t \sqrt{\frac{8\pi F}{a}}.$$

[We might have deduced this result from the expression for the energy contained in unit volume of the dielectric (§ 293), which is

$$\frac{1}{8\pi} F^2 = \frac{1}{8\pi} \frac{(V - V')^2}{t^2}.$$

Hence the total energy contained in the volume at between the discs is

$$\frac{1}{8\pi} \cdot \frac{(V-V')^2}{t} \cdot a.$$

And the rate at which this varies per unit of thickness gives the force

$$F = \frac{1}{8\pi} \cdot \frac{(V-V')^2}{t^2} \cdot a.]$$

In Kelvin's *absolute electrometer*, in which this principle is used, a concentric circular portion of the upper disc is alone moveable, and so the difficulty of the non-uniformity of the force at the edge of the discs is avoided. The moveable part as nearly as possible fills the aperture without touching its sides, and it is suspended by means of a delicate (spring) balance, which has a fiducial mark, by means of which the lower face of the moveable disc can always be placed in one plane with the lower face of the surrounding portion of the upper disc—called the *guard-ring*. The lower disc can be accurately moved, perpendicular to its own plane, through known distances, by means of a screw. The balance and disc are surrounded by a metal case for the purpose of preventing any disturbance which might arise from external electrification. When the two discs are connected to bodies of different potential, the balance is depressed, and the screw is turned until it returns to its standard position. In this way the distance t is determined. Also, by previous experiments, it is known what weight must be placed on the disc in order to bring it to its standard position, and this gives the (constant) value of F.

In using the instrument it is preferable to keep its lower disc at a constant potential by means of a charged condenser—the constancy being determined by means of a secondary electrometer. The value of t is then found, first, when the upper disc is connected to the ground, and, again, when it is connected to the body whose potential is to be determined. If θ be the difference of the distances, the above equation gives us the value of the potential

$$V = \theta \sqrt{\frac{8\pi F}{a}},$$

a being now the area of the moveable part of the disc.

In Kelvin's *quadrant electrometer* an aluminium needle swings inside a hollow metal cylinder, which is divided into four quadrants. The two opposite pairs of quadrants (Fig. 155) are connected by wires. The needle in its normal position is suspended with its length directed along one of the lines of division between the quadrants, and it is charged to a high positive potential. One pair of quadrants, say those connected to the wire *a*, may be connected to the ground, while the other pair is connected by the wire *b* to a body at positive potential V. The quadrants connected with *b* become positively charged, and the quadrants connected with *a* become negatively charged. The needle is therefore deflected towards the negatively charged quadrants, and the deflecting couple is proportional to V, if the potential of the needle be sufficiently high.

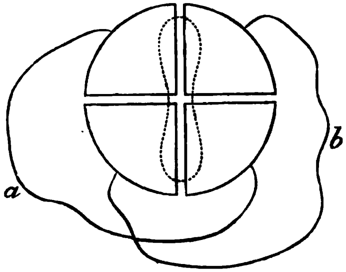


FIG. 155.

Modifications of this instrument may be used for the measurement of extremely small, and of extremely large, differences of potential.

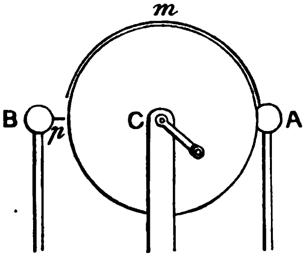


FIG. 156.

299. *Electric Machines.*—The electrophorus, which is the simplest form of electrical machine, has been already described.

As an example of the older machines used for the continuous production of electricity, we shall take the *cylinder machine.* This machine consists of a glass cylinder C (Fig. 156), which is turned round in the direction A*m*B. An insulated leather rubber A, coated with zinc amalgam, presses against the rotating cylinder, and causes the development of positive electricity on the glass, while it becomes itself negatively electrified. The positive electricity is carried round on the surface of the glass until it reaches the sharp metal points *p*, which project from the insulated metallic conductor B. It induces in the points negative electricity, which is discharged on the surface of the glass, destroying its electrification, and leaving B positively electrified. A silk flap *m*, connected to the rubber and resting in contact with the upper portion of the cylinder, prevents the electrification from slipping back along the surface of the glass. The potential of the positive electricity rises rapidly as it is carried from A to B, and the resulting electromotive force may cause the electricity to slip back, so that the potential of B cannot rise very high. The silk flap becomes negatively electrified, and so prevents the slipping, or stops it if it does occur by using the slipping electricity to destroy its own negative electrification, which tends to equalise the potential.

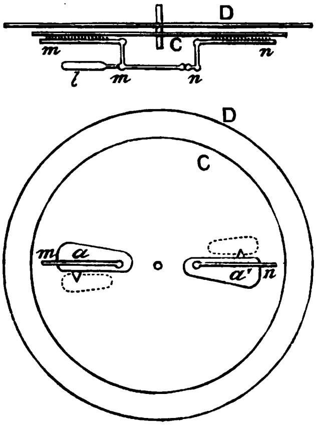


FIG. 157.

In the *plate machine* the glass cylinder is replaced by a circular glass plate, which is rubbed on both sides.

The *Holtz machine* is one of the best modern forms of electric machines. A fixed glass disc D (Fig. 157) has two paper armatures (*a*, *a'*) fixed on it near the opposite extremities of a diameter. Near each armature an opening (shown by dotted lines) is cut in the glass, through which a paper point attached to the armature projects so as nearly to come in contact with a revolving glass disc C, which is mounted upon an axis passing through the centre of D. A metal conductor *m*, fitted with a row of sharp points, faces the revolving disc on the other side opposite the armature *a*. A similar conductor *n* faces the armature *a′*, and can be placed in communication with *m* by means of the rod *l*, which slides through the knob *m*.

In order to work the machine, the knobs *n* and *m* are placed in connection, and a charge (positive, say) is given to the armature *a* by means of the electrophorus or otherwise, while the disc is rotated in the direction *aCa′*. After a short time a rustling sound is heard, and the machine becomes difficult to drive. It is producing electricity, and the extra work which has to be performed is the equivalent of the electrical energy which is developed.

We may explain the process of charging in the following manner: The positive charge given to *a* induces negative electrification in the points of the conductor *m*. This is discharged upon the glass surface, and the glass carries it round to the opposite side of the machine, leaving *m* positively charged. Here it induces positive electricity in the sharp point of the armature *a′*, and this is discharged upon the inside of the glass disc, leaving *a′* negatively charged. The armature *a′* now draws positive electricity to the points of the conductor *n*. This electricity is discharged upon the outer surface of the disc, leaving the conductor *n* negatively charged. Thus we may regard the glass disc as constantly carrying positive electricity from *n* to *m* in the one half of its revolution, and as constantly carrying negative electricity from *m* to *n* in the other half of its revolution.

The negative charge is also given to *a′* by way of the conductor *n*. The positive electricity, which is produced in *m* by the first motion of the machine, and flows to *n*, acts inductively on the armature *a′*, drawing negative electricity to the body of it, and repelling positive electricity to the sharp point to be discharged upon the glass, leaving *a′* negatively charged.

When the conductors *n* and *m* are slightly separated, a brush discharge passes through the air space. This brush discharge may be changed into a violent spark discharge by connecting the inner coatings of two Leyden jars to the conductors *n* and *m*, the outer coatings of the jars being joined together. The jars have to be charged up to the potential required for discharge through the air space, and, as their capacities are large, a great quantity of electricity passes at each discharge.

In more recent types of the Holtz machine, the electricity is not carried all the way from m to n, or from n to m, on the glass. The armatures are relatively much larger than those shown in the diagram. They extend over a range of about 45° on either side of each of the conductors m and n, and another metallic conductor, also armed with sharp points, is placed diagonally across the disc so that one row of points is opposite the upper edge of the armature a, while the other row is opposite the lower edge of the armature a'. If the discharging circuit mn be kept open, the action goes on, as first described above, by means of the diagonal conductor. The electricity which is discharged from the points of the horizontal conductor destroys the electrification of the glass, and there is thus a range of 45° on each side of the disc, throughout which the outer surface of the glass is not electrified.

Double Holtz machines are frequently used.

Self-exciting machines may be obtained by using a number of metallic sectors on a revolving disc and employing the principle of contact electrification of dissimilar metals.

The *Voss* machine is of this type. As in the Holtz machine, there is a fixed glass disc with two paper quadrants fixed to its remote side. To aid the action of each quadrant, tinfoil discs are attached to the upper and lower edges, and are joined by a strip of tinfoil. Tinfoil sectors are placed symmetrically on another glass disc which revolves in front of the fixed disc. To each sector a brass button is attached. A metallic conductor is fixed to the upper edge of one of the paper quadrants, and another is fixed to the lower edge of the other quadrant. Each carries a brush of fine metallic wire which touches the brass button of each sector as it passes. The discharging circuit is horizontal, and is also armed with brushes which touch the revolving sectors. A diagonal conductor, similarly armed, is arranged as in the Holtz machine above described. The action is essentially the same as that of the Holtz machine, with the exception that there is no discharge from sharp points, and that the initial electrification by contact of dissimilar metals is sufficient to make the machine self-exciting.

The machine is often doubled by employing two revolving discs.

In the *Wimshurst machine* there are two equal glass discs similarly provided with a large number of metallic sectors on one side. These discs revolve in opposite directions. The discharging circuit, armed with metallic brushes, is arranged horizontally. Diagonal conductors, one for each disc, are set, at an angle of 45° with the discharging conductor, so as to have their lengths perpendicular to each other.

It will be observed that machines of the Holtz, Voss, or Wimshurst, types, perform a double function. They multiply a charge which already exists in an armature, and they produce a supply of electricity for other purposes.

It is desirable sometimes to have an instrument which will perform the function of a multiplier alone. As an example, we may consider Lord Kelvin's *Replenisher* used for the purpose of varying the charge of the needle of the quadrant electrometer. In this apparatus there are two metallic inductors, segments of cylinders, which are oppositely charged. These face each other, and two insulated metallic carriers revolve, at opposite extremities of an arm, so as to pass in and out of them. By means of springs, attached to the ends of a metal rod, these carriers are placed in connection with each other when they are inside the inductors. They thus become oppositely charged. Each then passes out of the one inductor and is put, by means of a spring, in metallic connection with the other, at the instant that it enters it, and gives most of its charge to it. In this way the charges of the inductors may be increased until leakage becomes too great or disruptive discharge occurs. If the carriers be revolved in the opposite direction, the charges are diminished.

CHAPTER XXV.

THERMO-ELECTRICITY.

300. *Thermo-electric Phenomena.*—Though the principle of conservation of energy shows (§ 297) that the sum of the electromotive forces in a closed metallic circuit must be zero, provided that there be no difference of temperature between the various parts of the circuit, we cannot assert that their sum will be zero if the temperature be not uniform. For, in the process of equalisation of temperature, it is possible that there may be transformation of thermal energy into electric energy; and this transformation will occur if the electromotive forces of contact between the metals which form the circuit depend upon the temperature.

Now, Seebeck discovered in 1822 that, in general, a current of electricity flows around a circuit which is composed of two different metals, if there be a difference of temperature between the two junctions of the metals; and this shows that the equilibrium of the contact forces has been destroyed because of the variation of temperature.

And we cannot assert, without experimental evidence, that there will be no resultant electromotive force in a uniform circuit composed of a single metal which varies in temperature from point to point. But the experiments of Magnus showed that no such force exists.

Still, in order that Magnus's result should hold, it is necessary that every part of the circuit should be *physically* similar and not merely chemically similar. For example, two portions of the same metallic substance, which are in a different state of strain, are physically different substances, and are also thermo-electrically distinct. And two portions of the same substance which are at finitely different temperatures are in different physical states, and might, therefore, exhibit thermo-electric phenomena. Such phenomena were observed by Le Roux and others at the instant when contact was made between two portions of the same metal which differed abruptly in temperature.

301. *Laws of Thermo-electric Circuits.*—It is found by experiment that *the introduction of a piece of metal into a thermo-electric circuit does not contribute to the electromotive force of that circuit, provided that the extremities of the metal are at one and the same temperature.* We may, therefore, use solder to connect together the various parts of the circuit.

Let the lines A (Fig. 158) represent two pieces of the same metal. Let two of their ends have a common temperature t_1, while the other two have a common temperature t_0; and let the ends which are the temperature t_1 be joined by a metal C, while the ends which are at the temperature t_0 are joined by a metal B. As the whole arrangement is symmetrical with respect to the pieces A, it is obvious that there can be no resultant electromotive force in the circuit. And if the temperature of one of the junctions between C and A be changed from t_1 to t_2, the metal B will still, from its symmetrical position, contribute nothing to the electromotive force, although there may now be a resultant electromotive force in the circuit.

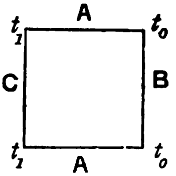


Fig. 158.

Next let us arrange pieces of two metals alternately, as in Fig. 159, and let the temperatures of their extremities be as indicated. The pieces B, which are at temperatures t_1 and t_2 respectively, contribute nothing to the total effect, so that the whole arrangement really consists of two metals whose junctions are at temperatures t_3 and t_0 respectively. Now, we may join the pieces t_1Bt_1 and t_2Bt_2 respectively to those points of the piece t_3Bt_0, which are at the temperatures t_1 and t_2, by means of connections made of the metal B; and these new connections contribute nothing to the total electromotive force in the circuit. But the electromotive force in the part BBt_0At_1B is due to the metals B and A when the temperatures of the junctions are t_1 and t_0. Similarly, the force in the part BBt_1At_2B is due to A and B, with temperatures t_1 and t_2 at the junctions; and that in the part BBt_2At_3 is due to A and B with temperature t_2 and t_3 at the junctions. Hence the electromotive force, due to temperatures t_3 and t_0, is equal to the algebraic sum of the electromotive forces due to temperatures t_1 and t_0, t_2 and t_1, t_3 and t_2 respectively.

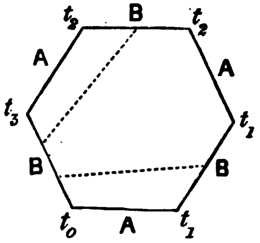


Fig. 159.

This result is quite general, and, therefore, *the algebraic sum of the various electromotive forces in a compound circuit, which is composed of a number of pieces of two metals with their junctions at various temperatures, t_0 and t_1, t_1 and t_2, ... t_{n-1} and t_n, is equal to the electromotive force in a two-junctioned circuit of the same metals with its junctions at the extreme temperatures t_0 and t_n.*

We can thus obtain a comparatively large electromotive force by means of a small difference of temperature. This is the essential principle of the *Thermopile*, an instrument which, in conjunction with a galvanometer, is used for the measurement of small differences of temperature (§ 307).

Lastly, arrange four metal wires, A, B, C, and B in the manner

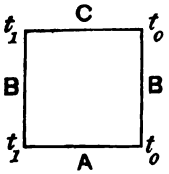


Fig. 160.

indicated in Fig. 160, and let the one wire B be raised to a temperature t_1, while the other is kept at a temperature t_0. The wires B merely serve as junctions, and so the total electromotive force is due to a circuit composed of A and C with the junctions at temperatures t_1 and t_0 respectively. But we may join the two wires B by a third wire of the same metal without altering the distribution of electromotive force in the circuit. And in the circuit Bt_1At_0BB the electromotive force is due to A and B with junctions at temperatures t_1 and t_0; while in the circuit Bt_0Ct_1BB the electromotive force is due to B and C, with junctions at the same temperatures.

If we define the *thermo-electric power* of a circuit of two metals as the rate at which the electromotive force in that circuit varies per unit of difference in temperature between the junctions, we see that the result which we have just obtained shows that *at every temperature the thermo-electric power of* A *and* C *is the algebraic sum of the thermo-electric powers of* A *and* B, *and* B *and* C.

302. *Variation of the Electromotive Force with Temperature.*—Soon after Seebeck's discovery, Cumming observed that in certain circuits (such as that of iron and copper), while one junction is maintained at a constant ordinary temperature and the other is gradually raised in temperature, the electromotive force gradually increases to a maximum, then diminishes, vanishes, and finally is reversed.

The law of variation of the electromotive force has been very fully investigated by Gaugain and others. They found that, with most pairs of metals, the curve which is obtained by plotting difference of temperature along the (horizontal) axis of x, and electromotive force along the (vertical) axis of y (Fig. 161) is in general a parabola with its axis vertical. Therefore, if we denote the electromotive force by e, and the temperature by t, and let E and T respectively represent the electromotive force and the temperature which correspond to the vertex of the parabola, we obtain

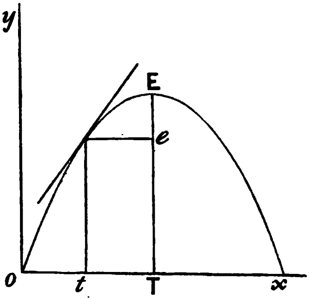


FIG. 161.

$$E - e = b(T - t)^2 \quad . \quad . \quad . \quad . \quad (1),$$

where b is a constant; for this equation merely expresses the well-known property of the parabola that the square of the distance of any point on it from the axis is proportional to the distance of that point from the tangent at the vertex.

In particular cases the curve is a straight line; in others, it is made up of portions of parabolas with parallel (vertical) axes, but with their vertices alternately turned in opposite directions.

303. *The Thermo-electric Diagram.*—Now another well-known property of the parabola is that the rate of increase of the ordinate at any point per unit of increase of the abscissa is proportional to the value of the abscissa measured from the axis. The proof of this is simple, for, if e_1, e_2, be the values of e at the temperatures t_1, t_2, respectively, (1) gives

$$\frac{e_2-e_1}{t_2-t_1}=b(2\mathrm{T}-t_2-t_1),$$

and, when t_2 and t_1 are indefinitely nearly equal to t—one being greater, the other less, than it—this becomes

$$\frac{e_2-e_1}{t_2-t_1}=2b(\mathrm{T}-t) \quad . \quad . \quad . \quad . \quad (2),$$

which is the direct expression of the above statement.

If, therefore, we plot the value of $(e_2-e_1)/(t_2-t_1)$—which is the thermo-electric power—against difference of temperature, we shall obtain, instead of a parabola, a straight line. And we may form a *self-consistent* diagram of such lines for any number of metals by means of observations on circuits consisting of each of these, in turn, with some standard metal whose line is made to coincide with the axis of temperature. (In the *true* diagram the lines might, of course, be curves obtained from these straight lines by a process of shearing.) We see, by (2), that these various lines will intersect the axis of temperature at points which correspond to the temperatures of the maximum ordinates in the original diagram (Fig. 161). And, further, the point of intersection of any pair of lines in the diagram indicates the temperature, T, at which the electro-

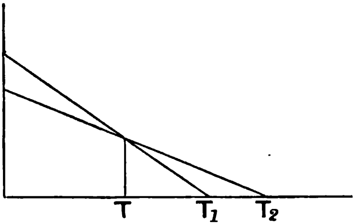


FIG. 162.

motive force in a circuit of the two corresponding metals attains a maximum value. This diagram is called the *thermo-electric diagram.*

Lord Kelvin first suggested the construction of such a diagram. The actual construction of it, upon the assumption (suggested by theoretical considerations) that the curves which represent the thermo-electric powers of the metals are, in general, straight lines, is due to Tait. The diagram on page 441 is reduced from his results.

The area included between two temperature lines and the lines which represent the thermo-electric position of any pair of metals represents the electromotive force in a circuit of these metals when the junctions are kept at the two given temperatures. This follows from the way in which the diagram of lines has been deduced from the diagram of parabolas.

304. *The Peltier Effect.* — A thermo-electric circuit forms a system which is in stable equilibrium. If it were in unstable equilibrium, increase of temperature of one of the junctions would produce effects which would still further increase the temperature. But we know that the application of heat to one junction produces a current of electricity which flows in a certain direction across that junction. Therefore, by the principle of stable equilibrium (§ 15), we can assert that the passage of electricity in the given direction will cool the junction.

The current at the hot junction always flows from the metal which has the lower thermo-electric power to the metal which has the higher thermo-electric power. Conversely, heat is absorbed at a junction where a current flows in this direction, or is evolved at a junction where the current flows in the reverse direction. Peltier discovered this by direct experiments made without reference to any theoretical considerations; and so the phenomenon of the absorption or disengagement of heat at a junction across which electricity flows is known as the *Peltier Effect* at that junction. The total Peltier effect in any circuit vanishes when the two junctions are at the same temperature, for the absorption of heat at one of the junctions is equal to the disengagement of heat at the other.

305. *The Thomson Effect.*—In order to explain the fact that, in such circuits as iron-copper, the direction of the electromotive force changes when the hot junction is sufficiently raised in temperature, Lord Kelvin (Sir W. Thomson) assumed that the Peltier effect vanishes at that temperature at which the electromotive force reaches its maximum value, that is, at the temperature at which the lines of the metals intersect in the thermo-electric diagram.

The metals are then said to be neutral to each other, and so this temperature is called the *Neutral Temperature.*

Now no heat is being absorbed or developed at the junction which is at the neutral temperature, and heat is being developed at the cold junction, for there the current is flowing from the metal of higher thermo-electric power to the metal of lower thermo-electric power. It would seem, therefore, that there is no source of thermal energy in the circuit by the transformation of which we can account for the development of electrical energy. But there is no other possible source of the electrical energy, and hence Kelvin was led to predict that heat is absorbed at parts of the circuit other than the junctions, either in that metal in which the current flows from hot parts to cold parts, or in that metal in which it flows from cold parts to hot parts, or in both metals. And he subsequently verified his prediction by direct experiment.

In copper, heat is absorbed when the current is passing from cold parts to hot parts; in iron, it is absorbed when the current is passing from hot parts to cold parts. [It is assumed here that we know the direction in which a current is flowing. The convention by which this is determined will be stated in next chapter.] Now, in a tube through which a liquid is flowing, heat is absorbed by the liquid when it passes from cold parts to hot parts. Hence, in copper and similar metals, electricity acts as an ordinary fluid would do; and so Kelvin speaks of the *specific heat* of electricity. It is positive in copper and similar metals, and is negative (at ordinary temperatures, at least) in iron and similar metals.

306. *Further Discussion of the Thermo-electric Diagram.*—The Peltier and the Thomson effects can be readily represented on the thermo-electric diagram.

Let e_1 represent the electromotive force in a certain circuit composed of some metal together with the standard one. T_1 being the neutral temperature, we get, at temperatures t and t' respectively, while the temperature t_0 of the cold junction remains constant,

$$E_1 - e_1 = b_1(T_1 - t)^2$$
$$E_1 - e'_1 = b_1(T_1 - t')^2$$

whence

$$e'_1 - e_1 = 2b_1(t' - t)\left(T_1 - \frac{t'+t}{2}\right) \quad . \; . \; . \; . \; (3).$$

Similarly, if we use any other metal giving the electromotive force e_2 with the standard metal under the same conditions,

$$e'_2 - e_2 = 2b_2(t' - t)\left(T_2 - \frac{t'+t}{2}\right) \quad . \; . \; . \; . \; (4).$$

Hence the electromotive force in a circuit of the two given metals under the same conditions as to temperature is

$$\begin{aligned}\bar{e}&=(e'_1-e_1)-(e'_2-e_2)\\&=2(t'-t)\left[b_1\mathrm{T}_1-b_2\mathrm{T}_2-(b_1-b_2)\frac{t'+t}{2}\right].\end{aligned}$$

Now (2) shows us that $2b_1$ and $2b_2$ are the tangents of the angles at which the lines of the metals meet the line of the standard metal in

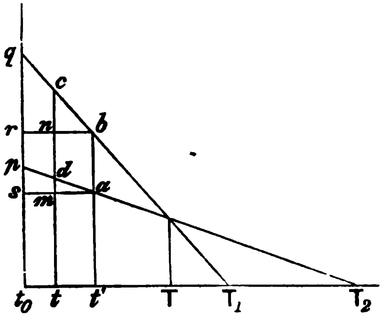


FIG. 163.

the thermo-electric diagram. So, if we let t_0 represent the absolute zero of temperature, while all the other temperatures are given in absolute units, we get (Fig. 163)

$$qt_0=2b_1\mathrm{T}_1 \text{ and } pt_0=2b_2\mathrm{T}_2,$$

and so $\quad qp=2(b_1\mathrm{T}_1-b_2\mathrm{T}_2)$. But $qp=2\mathrm{T}(b_1-b_2)$,

whence $\quad b_1\mathrm{T}_1-b_2\mathrm{T}_2=\mathrm{T}(b_1-b_2) \quad . \; . \; . \; . \; (5)$

and therefore $\quad \bar{e}=2(b_1-b_2)\,(t'-t)\left(\mathrm{T}-\dfrac{t'+t}{2}\right) \quad . \; . \; . \; . \; (6),$

which is the general expression for the electromotive force in a circuit of two metals in terms of the inclinations of their lines to the line of the standard metal, of their neutral temperature, and of the temperatures of the junctions.

Now (6) may be put in the form

$$\bar{e}=2(b_1-b_2)(\mathrm{T}-t')(t'-t)+(b_1-b_2)(t'-t)^2 \quad . \; . \; . \; . \; (7).$$

The first term on the right-hand side of (7) vanishes when $t'=t$, and also when the temperature of the hot junction is at the neutral point. It therefore represents the part of the electromotive force which corresponds to the Peltier effect. Therefore, if there is no other effect than the Peltier effect and the Thomson effect, the second term must represent the part of the electromotive force which corresponds to the Thomson effect.

If we suppose that unit quantity of electricity is transferred along the circuit under the electromotive force $\bar{e}$, the quantity $\bar{e}$ represents the electric energy which is expended in the process, and therefore the quantity on the right-hand side of the equation expresses, on the one hand, the heat which is absorbed in the production of the electric energy, or, on the other hand, the heat which is evolved, in the manner under consideration, when unit quantity of electricity passes round the circuit, in the reverse direction, under the electromotive force $\bar{e}$. The first term of (6) is therefore taken as the measure of the Peltier effect, while the second term is taken as the measure of the Thomson effect.

If the current flows round the circuit in the direction *abcda*, the area *abcda* represents the whole heat which is absorbed in the circuit during the passage of unit quantity of electricity.

But $2(b_1-b_2)(\mathrm{T}-t')=ab$, and therefore the area *abnma* represents the heat which is on the whole absorbed at the junctions. The whole area *abrsa* represents the heat which is *absorbed* at the hot junction, for at that junction the current is passing from the metal of lower thermo-electric power to the metal of higher thermo-electric power; while the area *msrnm* represents the heat which is *evolved* at the cold junction, for at the cold junction the current passes from the metal of high thermo-electric power to the metal of low thermo-electric power.

Again, $cn=2b_1(t'-t)$, and therefore the triangular area $cnb= \frac{1}{2}\cdot 2b_1(t'-t)(t'-t)$ represents the heat which is *absorbed* in the metal of higher thermo-electric power. Similarly the area *amd* represents the heat which is *evolved* in the metal of lower thermo-electric power. [The former part has the positive sign prefixed to it in (7); the latter part has the negative sign prefixed.]

It is evident that we may state quite generally that heat is absorbed at any part of the circuit at which the current is passing from parts of lower thermo-electric power to parts of higher thermo-electric power, and is evolved at any part when the current is passing from parts of higher to parts of lower thermo-electric power.

The term $(b_1-b_2)(t'-t)(t'-t)$ may be regarded as the product of the sum of the average specific heats of electricity in the various parts of the circuit into the range of temperature, each term in the sum being positive or negative according as it corresponds to absorption or evolution of heat. Therefore the form of the various terms b_1t', b_1t, b_2t', and b_2t, shows that the specific heat of electricity in any metal is proportional to the absolute temperature, and

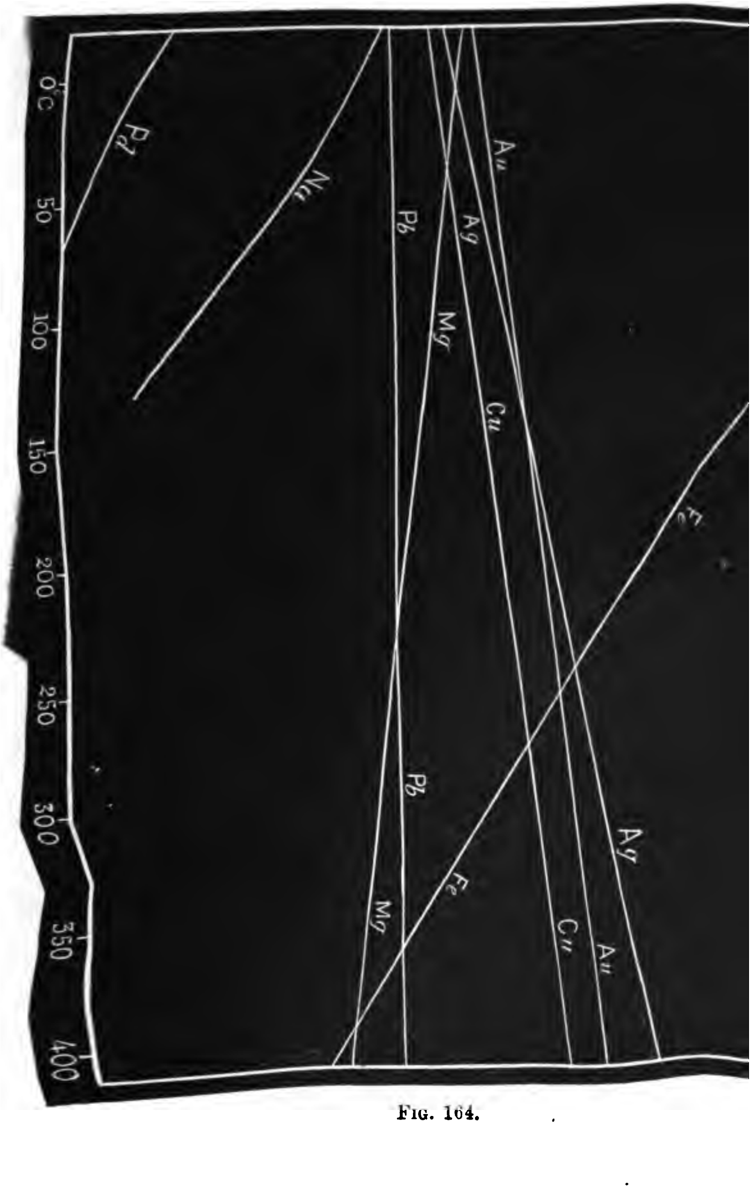

0°C
50
100
150
200
250
300
350
400
Pd
Ni
Pb
Ag
Au
Mg
Cu
Fe
Pb
Ag
Fe
Mg
Cu
Au

FIG. 164.

consequently the average which must be taken is the arithmetical mean; so that, if σ be the actual specific heat at temperature t, the average value throughout the range t will be $\frac{1}{2}\sigma$. Thus, when the electricity flows from b to c (Fig. 163), we may suppose that it flows along bq, and comes back along qc. In the first part of this process heat is absorbed; in the second part heat is evolved. In the first part the average specific heat is (say) $\frac{1}{2}k_1t'$, where k_1 is a constant. Similarly, in the second part, the average specific heat is $\frac{1}{2}k_1t$, which must be regarded as a negative quantity, since heat is being *evolved* as the electricity passes from cold parts to hot parts. The total sum for the two metals is therefore $\frac{1}{2}(k_1-k_2)\ (t'-t)=\sigma_1-\sigma_2$, so that the whole product $(b_1-b_2)\ (t'-t)\ (t'-t)$ is identical with $(\sigma_1-\sigma_2)\ (t'-t)$, and k_1 and k_2 are double of b_1 and b_2 respectively.

The specific heat of electricity in the metal whose line is $q\mathrm{T}_1$ (Fig. 163) is therefore represented by the line qr at the temperature t', and so on. It is negative or positive, according as the line slopes downwards (like that of iron, Fig. 164) or upwards (like that of copper).

Le Roux found that the specific heat of electricity in lead is zero (or very nearly so), and therefore lead is chosen as the standard metal in the construction of the diagram.

Tait has found the very curious result that the specific heat of electricity in paramagnetic metals, such as iron and nickel, changes sign at least twice as the temperature is raised.

The following table contains some of Tait's results. The second column gives the value, in microvolts, of the thermo-electric power relatively to lead at 0° C. The third column gives the correction to be applied per degree Centigrade when the temperature differs from 0° C.:

Iron ...	17·34	−0·0494
Cadmium	2·66	0·0436
Platinum (soft)	2·60	−0·0112
Zinc ...	2·34	0·0244
Silver ...	2·14	0·0152
Copper ...	1·36	0·0096
Lead ...	0	0
Tin ...	−0·43	0·0056
Aluminium	−0·77	0·0040
Palladium	−6·25	−0·0364
German silver	−12·07	−0·0512
Nickel ...	−22·04	−0·0520

The temperature correction must not be applied in the case of nickel at a temperature above 175° C. The correction given for iron may be used until the temperature of dull redness is reached. The neutral points with lead may be obtained from these data by dividing the numbers in the second column by the corresponding numbers in the third column with the sign changed. And, since the number in the third column, corresponding to a given metal, represents the value of $2b$ for that metal, the neutral point of any pair of the given metals may be calculated by means of the formula (5).

307. *The Thermopile and Radiomicrometer.*—Two conditions are essential in the construction of an efficient thermopile which is intended for the production of strong electric currents. Conduction of heat from a hot junction to a cold junction must be avoided, and the resistance to the passage of the current must be as small as possible; and these conditions are mutually inconsistent. Hence the thermopile can never replace the galvanic cell in its ordinary uses.

The chief defect of the thermopile, when it is employed for the detection or measurement of radiation, is its sluggishness. A considerable mass of metal has to be heated, and so the current takes a considerable time to rise to its full value.

This difficulty has been overcome by C. V. Boys in his *radiomicrometer*. To one end of a small loop of copper wire, a light bar of bismuth is soldered, while a light bar of antimony is soldered to the other end. The circuit is closed by soldering these bars to a disc, or a narrow strip, of blackened copper foil on which the radiation falls. The whole arrangement is suspended, by a quartz fibre, in the field of a powerful magnet.

The same principle was employed previously by D'Arsonval.

CHAPTER XXVI.

ELECTRIC CURRENTS.

308. *Convection Current between Charged Conductors.*—A pith-ball placed between two oppositely charged insulated conductors, and free to move between them, will gradually destroy their electrification. When it comes in contact with the positively charged body, it receives a positive charge with which it moves, under the action of the electric force, towards the negatively charged body. It gives its positive charge to this body, and receives from it a negative charge, with which it again moves towards the positive conductor, and so on alternately.

If the charges of the two conductors were originally equal, the process will result in the complete destruction of their electrification. If only one of the two is originally charged, the process will result in the division of the charge between the two in the ratio of their capacities.

Positive electricity is carried in one direction, and negative electricity is carried in the reverse direction, by a convective process. The greater the electric force between the conductors is, the more rapid will be the motion of the ball, and the more nearly will the process approach to a continuous flow of electricity.

309. *Flow of Electricity in Metallic Conductors.*—If we place the two oppositely charged conductors in contact with each other by means of a metallic wire, a current of electricity will be established between them until their potentials are equalised. If one of the bodies be electrified, say positively, while the other is unelectrified, the result is (as above) that the charge is divided between the two bodies in the ratio of their capacities; and it is usual to say that positive electricity has flowed from the first body to the second, though the same result might have been produced by a flow of induced negative electricity in the opposite direction.

Any difference of potential between two parts of a conductor constitutes an electromotive force under which transference of positive electricity will take place from the part at higher potential to the part at lower potential. And so long as the potentials are maintained constant, the quantity of electricity which flows from the one part to the other in a fixed time remains constant. The amount which flows per unit of time along the conductor from the one part to the other, is called the *Strength* or *Intensity* of the current. The practical unit of current strength is called the *ampere*. (See the concluding sections of Chapter XXIX.)

The slightest difference of potential between two parts of a conductor will produce a current of electricity between these parts, but no current can be maintained without the expenditure of energy. That is to say, the flow of the current is opposed by a *Resistance* in the conductor. This is analogous to the flow of a liquid along a tube. A current is produced in the tube by the action of the slightest force, but work must be expended in order to maintain the flow; for the motion is opposed by a resistance which is due to friction.

The practical unit of electric resistance is called the *ohm*.

When an incompressible fluid flows through a tube because of a constant difference of pressure at its extremities, equal quantities of fluid pass every section of the tube in the same time. So, in the flow of electricity along a conductor because of a constant difference of potential at its extremities, equal quantities of electricity pass every section of the conductor in the same time.

When by any means a difference of potential is maintained at different parts of a conductor, we may draw in the conductor a series of equipotential surfaces. The electric stream lines are everywhere perpendicular to these surfaces.

Fig. 165 (*a*) represents the distribution of stream lines and equipotential lines in a thin circular conducting sheet, the centre of which is kept at a constant negative potential, while a point on its circumference is kept at an equal positive potential. The equipotential lines are in part open curves with their extremities on the circumference of the circular sheet, and in part they are closed curves surrounding the centre. If the whole diagram be revolved about the axis of symmetry, the various lines will trace out surfaces of flow and of constant potential in a conducting sphere, when the given points are maintained at constant (different) potentials.

In the upper half of Fig. 165 (*b*) the method by which the stream lines in the above case are drawn is exhibited. The circumference of the circle is divided into a whole number of equal parts, and lines are drawn from A and C to the extremities of these parts. These lines are numbered from A round the circumference to the opposite end of the diameter AC.

The lower half of the figure exhibits the method of drawing the equipotential lines when A and C are points in an infinitely extended conducting sheet. The nature of the difference between them and the equipotential lines of Fig. 165 (*a*) is apparent.

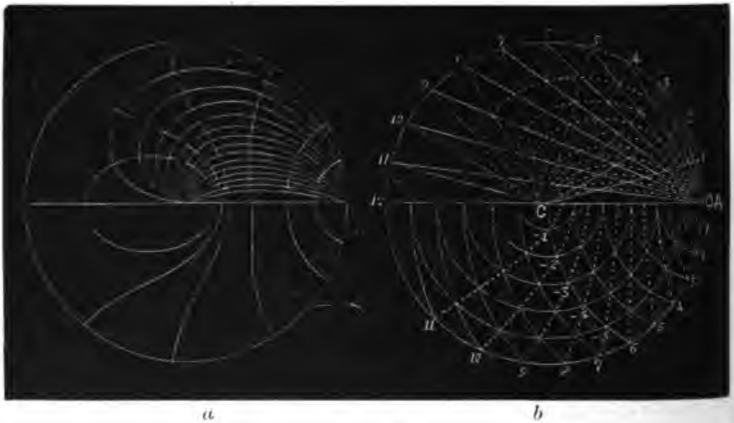


FIG. 165.

310. *Ohm's Law. Kirchoff's Laws.*—Ohm determined experimentally the relation which connects electromotive force, current-strength, and resistance in a conducting circuit. This relation is therefore known as Ohm's Law.

He found that in a conductor of given resistance the strength of the current is proportional to the electromotive force, and that in a conductor of variable resistance the strength of the current is inversely proportional to the magnitude of the resistance, if the electromotive force be constant. If E, C, and R represent respectively the electromotive force, the strength of the current, and the resistance, these results are expressed by the equation

$$E = CR$$

if we define unit current as the current which is maintained in a conductor of unit resistance by unit electromotive force.

This law enables us to calculate the resistance of a compound conductor composed of a number of conductors arranged in 'series,' that is, arranged so that the current flows from one to another of the several conductors in succession. Let E denote the total difference of potential in the circuit, and let R be the total resistance. Also, let e_1, e_2, etc., and r_1, r_2, etc., represent the corresponding quantities for the several conductors in the circuit. The same current C flows through all the conductors, and hence the condition

$$E = e_1 + e_2 + \ldots .$$

gives, by Ohm's Law,

$$R = r_1 + r_2 + \ldots . \quad (1).$$

Therefore *the resistance of a number of conductors arranged in series is the sum of their separate resistances.*

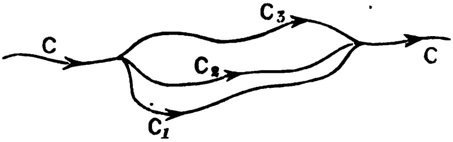


FIG. 166.

If the conductors be arranged in 'multiple arc,' as in Fig. 166, the condition

$$C = C_1 + C_2 + \ldots .$$

where C_1, C_2, etc., are the currents in the separate parts of the circuit, gives, by Ohm's Law,

$$\frac{E}{R} = \frac{E}{r_1} + \frac{E}{r_2} + \ldots . ,$$

whence

$$\frac{1}{R} = \frac{1}{r_1} + \frac{1}{r_2} + \ldots . \quad (2).$$

The reciprocal of the resistance of a conductor is called its *conductance*, and so this equation expresses the fact that the *conductance of a compound conductor, formed of a number of separate conductors joined in the multiple arc arrangement, is the sum of the individual conductances of these conductors.*

The practical unit of conductance is called the *mho*. Thus a resistance of R *ohms* has a conductance of 1/R *mhos*.

The law alluded to above, that the flow of electricity in conductors resembles that of an incompressible fluid, together with Ohm's Law, enables us to express the relations connecting electromotive force, current-strength, and resistance in the various parts of any network of conductors, however complex. These laws, stated in the forms,

(1) The sum of the currents which flow towards any point of the network is equal to the sum of the currents which flow from it,

(2) The sum of the electromotive forces which act in any closed loop of the network is equal to the sum of the products of the current into the resistance in the several parts of the loop,

are known as Kirchoff's Laws.

Conductivity and Resistivity.—From equation (1) we see that the resistance of a cylindrical wire of uniform conducting material is proportional to the length of the wire. And, if we regard the wire as composed of a number of wires, of unit section, laid side by side, equation (2) shows that the resistance is inversely proportional to the sectional area of the wire. Hence, if R be the total resistance, while l and s represent respectively the length and section, of the wire, we may write

$$R=\rho\frac{l}{s},$$

where ρ is a constant. If l and s be each equal to unity, this becomes $R=\rho$. So that ρ is the resistance of a unit cube of the material of the wire. This is called the *resistivity* or *specific resistance* of the material.

Similarly, if K represents the conductance of the wire, we get

$$K=\kappa\frac{s}{l},$$

where κ is a constant which is called the conductivity of the material.

These equations show that the resistance and the conductance are constant in the case of similar bodies, of a given material, under given conditions.

311. *Electrolytic Conduction. Faraday's Laws.*—When electricity passes through certain conductors, chiefly liquids, decomposition of the conductor takes place; and it appears that the decomposition is a *necessary* accompaniment of the passage of the electricity. Such substances are called *electrolytes*, and the process of decomposition is termed *electrolysis*.

The current usually enters and leaves the electrolyte by metallic conductors, which are termed the *electrodes*, the conductor by which the current enters being distinguished as the *anode*, while the conductor by which it leaves is called the *cathode*.

The products of decomposition appear at the electrodes, the metallic constituent appearing (unless some secondary chemical action prevents this) at the cathode, while the other constituent appears at the anode. Thus, in the electrolytic decomposition of hydrochloric acid, hydrogen appears at the cathode, while chlorine is evolved at the anode. No decomposition occurs in the interior of the electrolyte. Hence it follows that when an electromotive force acts upon the electrolyte, the metallic constituent travels in the direction in which the current goes (that is, from high to low potential), and the other constituent travels in the opposite direction. The two constituents are therefore termed *ions*, the one which moves towards the cathode being called the *cation*, while the one which moves towards the anode is called the *anion*.

The laws of electrolytic conduction were fully investigated by Faraday. He found that the amount of the electrolyte which is decomposed by the passage of a certain quantity of electricity is strictly proportional to that quantity. The amount which is decomposed during the passage of a unit of electricity is called the *Electrochemical Equivalent* of the substance.

Faraday also found that the amount of any substance, which appears as anion or cation, is totally independent of the substance with which it is combined; and this shows that *the electrochemical equivalent of a substance is absolutely constant.*

A current, whose strength is one ampere, deposits per second 0·000010384 gramme of hydrogen. This quantity is therefore taken as the electrochemical equivalent of hydrogen. If it be multiplied by the atomic weight and the valency of any other element, the result is the electrochemical equivalent of that element.

The process of electrolysis has received extensive practical applications in the art of electrometallurgy.

Instruments used to measure quantity of electricity by amount of chemical decomposition are termed *Voltameters*.

312. *Polarisation. Von Helmholtz's Electrochemical Theory.*—The passage of electricity through a liquid results in the chemical decomposition of that liquid, and the electrical energy is transformed in part into potential energy of chemical separation.

Now, if two parts of a conductor are kept at constant (different) potentials, the work which is done in transferring unit quantity of electricity from the part at the higher potential to the part at the lower potential is equal to the difference of potential between the two parts—that is, to the electromotive force (E, say). Let H be the amount of heat which is developed in the combination of unit amount of the ions, from the state in which they are liberated, so as to form the compound electrolyte. This amount of heat is equivalent to the amount of work JH, where J is the dynamical equivalent of heat. Hence, in dynamical units, JqH is the work which is expended when unit amount of electricity passes through an electrolyte of which q is the electrochemical equivalent, and, therefore,

$$E=JqH.$$

To reduce this electromotive force, in any particular case, to its value expressed in terms of the practical unit of electromotive force, it must be divided by the number of absolute units contained in the practical unit (Chap. XXIX., last section).

The quantities, J, q, and H, in this equation are all finite, and so E is finite; and, therefore, a finite electromotive force is required, in order to effect electrolytic decomposition. And, in a mixture of electrolytes, for each of which the value of E is different, that one for which E has the least value will electrolyse more readily than the others.

If we keep the electrodes at an insufficient difference of potential no decomposition can ensue; but if there be any difference of potential in the circuit, flow of electricity must occur. And we are here met with a difficulty, for Faraday's Law asserts that decomposition takes place in strict proportion to the amount of electricity which passes. But this difficulty can readily be explained. The phenomenon is precisely analogous to the charging of a Leyden jar. When the jar is charged under given conditions regarding potential, positive electricity flows into the one coating and negative electricity flows into the other until the potential of the jar is equal to that of the machine which is used to charge it. But the electrical energy is stored up in the dielectric which insulates the coatings of the jar, and will produce an equal reverse current of electricity whenever the coatings are put in metallic communication with each other, and we say that this reverse current is produced by the reverse electromotive force of the jar. Similarly, when an electromotive force, too feeble to cause decomposition, acts upon an electrolyte, flow of electricity takes place in the conducting parts of the circuit —that is, in the metallic and electrolytic conductors. But no transference of electricity can take place between the electrode and the electrolyte, for decomposition must occur (by Faraday's law) if it did, and the electromotive force is not large enough to produce decomposition. The layer (of thickness comparable with molecular dimensions), between the molecules of the electrolyte and the molecules of the electrode acts as an insulator, and its two surfaces become oppositely charged, as those of a Leyden jar would, until the reverse electromotive force produced in this way becomes equal to the direct electromotive force. If now the direct force be removed and the electrodes be joined so as to form a closed circuit, the reverse force will produce a reverse current.

This phenomenon is called *polarisation*, the reverse force is called the electromotive force of polarisation, and the reverse current is called the polarisation current.

Let E_0 be the force which is required to produce decomposition. If E be the direct force, the difference $E - E_0$ is alone effective in producing a permanent current.

The quantity E_0 is not constant except under fixed conditions; and, in practice, the conditions under which the decomposition takes place vary greatly. The electrodes or the electrolyte may have some chemical or molecular attraction for the products of decomposition, and the electromotive force which is required to produce the decomposition will be less in proportion as this attraction is greater. As an extreme case, an infinitely small electromotive force would effect the decomposition of an electrolyte, the constituents of which had no greater attraction for each other than they had for the substance of the electrode. Again, if the products of electrolysis are gaseous, and are dissolved by the electrolyte, decomposition will take place more readily than it would if the gases were not dissolved. When the process is first started the gases may dissolve readily, but their solution will occur with greater and greater difficulty until saturation is reached, when it entirely ceases. Further, if the gases evaporate from the electrolyte, saturation will never be fully attained, and the maximum attainable state of saturation will depend upon the partial pressure of the gases in the atmosphere which is in contact with the electrolyte.

These and other causes which affect the electromotive force have been fully discussed by Von Helmholtz, who has given a complete thermodynamical theory of polarization, the results of which accord very well with experimental facts.

If n be the number of atoms in the electrochemical equivalent of a substance, the quantity q/n may be regarded as the 'atomic charge,' which is constant since q and n are constant. This constancy of the atomic charge points to an intimate relation between electricity and matter, and the fact that the electrochemical equivalent of a substance is constant whether it be electrolysed from combination with a monad, dyad, or triad, etc., element shows that the atomic charge of a dyad element is double that of a monad element, that the atomic charge of a triad element is three times that of a monad element, and so on. This led Von Helmholtz to regard chemical affinity as the result of a greater attraction of some atoms than of others for positive or negative electricity. Thus, for example, oxygen has a greater attraction for negative electricity than hydrogen has, while hydrogen attracts positive electricity more strongly than does oxygen. And Von Helmholtz regards the affinity of oxygen and hydrogen for each other as the result of the electrical attraction between positively charged hydrogen and negatively charged oxygen.

Substances are classed as electro-positive or electro-negative, according as, in this respect, they resemble hydrogen or oxygen. The more electro-positive a substance is, the less readily will it act as an anion under given conditions if this view be correct; for the more strongly will it be held in combination with the electro-negative component of the electrolyte. As far as the metals are concerned, the series so determined should roughly agree with the series as determined by contact (§ 297); but the exact order depends upon the temperature, the nature of the solution, and other conditions. The correspondence is really very close. The order is, in the former case, potassium, sodium, zinc, iron, hydrogen, lead, copper, silver, platinum, carbon. In the latter lead precedes iron.

The atomic charge is excessively small since n is very large; but, since the atoms are at an exceedingly small distance apart, the attraction may be very great. The value of the charge is probably about 10^{-20} of the practical unit of quantity.

These considerations enabled Helmholtz to explain a phenomenon which appeared to be at variance with Faraday's law. An extremely feeble constant current, inappreciable to all but the most delicate instruments, flows in an electrolytic circuit under electromotive forces far too feeble to produce decomposition. Von Helmholtz showed that this was due to the presence of dissolved gases. Thus oxygen in solution will gradually find its way to the negative electrode and will receive a negative charge. It then becomes subject to the electromotive forces and travels to the positive electrode, to which it gives up its negative charge. In this way a constant 'convective' current of electricity is kept up, and Faraday's law is not subject to exception.

Helmholtz observed the existence of this convective current even when he had carefully freed the electrolyte from dissolved gases by the most perfect processes. Here again there is no real exception to Faraday's law, for Helmholtz's thermodynamical theory shows that some dissolved gases must be present if the (liquid) electrolyte is to be in stable equilibrium; and, therefore, even if all dissolved gases were abstracted, decomposition of the liquid would go on until the requisite minimum of dissolved gaseous products were present.

313. *Ideas of Grotthus and Clausius.*—The views of Von Helmholtz, above considered, constitute a development of previously formulated ideas.

Grotthus suggested that, under the action of electromotive force, the molecules of an electrolyte, say HCl, arranged themselves in polarised chains, thus :

$$\mathrm{HCl\ HCl\ HCl\ HCl\ HCl};$$

and that, when decomposition occurred, the electro-positive constituent of a molecule next the cathode was set free, while the electro-negative constituent of a molecule next the anode was set free, and a rearrangement of the other constituents into molecules took place, thus :

$$\mathrm{H\ ClH\ ClH\ ClH\ ClH\ Cl}.$$

The electromotive force still acting, this became

$$\mathrm{H\ HCl\ HCl\ HCl\ HCl\ Cl},$$

and the process was repeated.

Clausius added to this the idea that dissociation was not directly produced by the acting electromotive force. The force, according to his view, only acted upon electrified atoms already dissociated by molecular collisions.

This view readily accounts for the fact that, under a feeble electromotive force, a direct current lasts until the reverse force of polarisation becomes equal to the direct force.

314. *Ohm's Law in Electrolytes.*—In the case of electrolytes, Ohm's law must be put into the form $E - E_0 = CR$, E_0 being, as above, the electromotive force of polarisation. Its applicability has been verified, in all cases, within the limits of experimental error. The variability of E_0 renders the proof of the law extremely difficult. The quantity R is also variable (see § 320).

315. *Production of Electric Currents.*—An electrolytic circuit constitutes a system which is in stable equilibrium, for a finite electromotive force is required in order that decomposition may take place, and the removal of the electromotive force causes the cessation of the decomposition. Consequently, we conclude (§ 15) that, in this stable system in which the passage of a direct current produces chemical separation, the energy of chemical separation will be transformed into the energy of a reverse electric current, when the conditions are such that this reverse current can flow.

Therefore, in a closed conducting circuit, which includes an electrolyte, and which has associated with it energy of chemical separation, the sum of the electromotive forces may not be zero, but may produce an electric current, while chemical combination proceeds. The sum of the forces can differ from zero only because there is a chemical source for the energy of the current which the resultant force produces; just as, in a thermo-electric circuit, the electric energy is derived from the heat energy which is supplied.

An arrangement of this kind may therefore be used for the production of an electric current; and it is of no consequence whether the chemical separation is produced by the previous passage of an electric current through the circuit under the action of an external electromotive force, or is given as an initial condition produced by chemical processes or otherwise.

An arrangement of the latter class constitutes a *Primary Cell;* one of the former class constitutes a *Secondary Cell.*

The equation in § 312 indicates a method of determining the electromotive force in any such circuit; for it asserts that 'the electromotive force of an electrochemical apparatus is in absolute measure equal to the mechanical equivalent of the chemical action on one electrochemical equivalent of the substance.' This was first pointed out by Lord Kelvin. Conversely, the energy which is developed in a given chemical action can be estimated by means of measurements of the electromotive force of an electrochemical apparatus in which that action is developed.

316. *Primary Cells.* — Primary cells are sub-divided into two classes according as one fluid or two fluids are used in their construction.

All the older cells were of the one-fluid type, and the conducting circuit was generally made up of a plate of zinc, a plate of copper, and sulphuric acid. The sulphuric acid, combining with the zinc, produces zinc sulphate, and hydrogen is liberated at the copper electrode. Positive electricity flows from the zinc to the copper through the liquid (Fig. 167).

Such a cell has many disadvantages. Slight impurities in the zinc (possibly even differences in its physical constitution) at different parts give rise to local currents of electricity between these

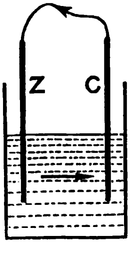


FIG. 167.

parts, and thus lead to solution of the zinc without the production of effective currents in the main circuit. This local action may be largely prevented by amalgamating the surface of the zinc, for the surface is thus rendered more nearly homogeneous. Again, the evolution of hydrogen at the copper electrode gives rise to a reverse electromotive force of polarization. And, ultimately, when the sulphuric acid is transformed into zinc sulphate, zinc, instead of hydrogen, is deposited upon the copper plate. When this process is complete, both electrodes are practically composed of zinc; and, from the symmetry of the arrangement, it is evident that the sum of the electromotive forces in the circuit must be zero, so that the current ceases.

Daniell improved this cell by the introduction of two fluids separated by a porous diaphragm which did not prevent the passage of electricity. A zinc rod Z (Fig. 168) is placed inside a porous cell, which is contained in an outer copper vessel C. The porous cell contains a solution of zinc sulphate, and the copper vessel contains a saturated solution of copper sulphate. In order to keep this solution saturated, crystals of the sulphate of copper may be placed on a perforated tray T, inside the copper vessel. These crystals are dissolved away when the solution becomes weakened.

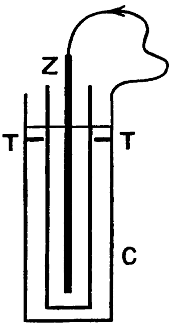


FIG. 168.

When the cell is in action, the current passes from zinc to copper through the liquid. The zinc sulphate is electrolysed into zinc and 'sulphion' (SO_4). The SO_4 acts upon the zinc and forms zinc sulphate. Thus the zinc sulphate always tends to be saturated. The copper sulphate is simultaneously electrolysed into copper and sulphion, and the copper (being the metallic ion, and therefore travelling with the current which is assumed to travel in the direction in which positive electricity goes) is deposited upon the copper vessel. The zinc, which results from the electrolysis of the zinc sulphate, is not deposited at all, but unites with the sulphion produced by the electrolysis of the sulphate of copper, and so re-forms zinc sulphate, which remains inside the porous cell. Thus hydrogen is not evolved at the copper electrode, and polarisation is reduced to a minimum.

If the current be too strong, or if the copper sulphate be not saturated, some hydrogen will be evolved.

Daniell's cell gives an extremely constant electromotive force if the conditions which are necessary to its proper action are maintained.

Bunsen's cell is also a two-fluid cell. A rod of carbon is placed in strong nitric acid, which is contained in the porous cell, and a zinc plate is placed in an outer cell of glazed earthenware which contains an aqueous solution of sulphuric acid. Hydrogen is evolved at the carbon, but is at once oxidised by the nitric acid. The fumes which are given off from the acid are very poisonous, so that these cells should be kept in a separate room, as far as possible, if a considerable number of them are in use.

Grove's cell resembles that of Bunsen in its chief features. The carbon is replaced by a sheet of platinised silver. The process of platinisation gives a large, very finely corrugated surface, which favours the diminution of polarisation.

The electromotive forces of the two latter cells are very nearly equal to each other, and are considerably larger than that of a Daniell cell, but are not so constant.

The single-fluid *bichromate cell* (bichromate of potassium dissolved in an aqueous solution of sulphuric acid into which dip zinc and carbon plates) gives a high electromotive force, and is greatly used when a powerful current is required for short periods.

The *Léclanché cell* consists of a porous cell, filled with manganese dioxide and powdered carbon, in which a carbon rod is placed. A zinc rod is placed in a solution of ammonium chloride, outside the porous cell. The electromotive force of this cell rapidly diminishes when it is made to produce a current, but it soon attains its original value after the current is stopped. The great advantage of the cell is that it remains in good working order for months.

The Daniell cell, because of its constant electromotive force, may be used as a standard for electromotive force. *Clark's standard cell* is even more constant. The cathode, pure mercury, is placed at the foot of the cell, and above it lies a paste of mercurous sulphate.

Above this there is a concentrated solution of zinc sulphate, into which the anode—a plate of pure zinc—dips. The whole is corked tightly, and cement, through which pass the wires from the electrodes, is filled in above the cork.

Many other forms of cells, including modifications of the above, are in practical use; but the present examples will serve sufficiently as illustrations of their general principles.

The zinc and the carbon, or the zinc and the copper, etc., are called the *elements* of the cell; carbon and copper being positive elements, while zinc is a negative element (for the current flows from carbon or copper to zinc through the external conductor).

Cells are sometimes divided into three classes, according to the method used to prevent polarisation. In one of these classes, polarisation is diminished by mechanical action, such as stirring the liquid. In another, a chemical *depolariser* is used. The bichromate cell is an example of this type—the bichromate of potassium oxidises the hydrogen which is liberated from the sulphuric acid solution. The Léclanché cell is another example of this type. In this case the depolariser is solid—manganese dioxide. In the third class, of which the Daniell cell and the Clark cell are examples, electrochemical depolarisation is employed. The cathode is made of the same metallic substance as the cation of the liquid into which it dips; and the product of the combination of the substance of the anode with the anion of the salt solution into which it dips is that salt itself. If that solution be saturated, and if the physical state of the metal liberated at the cathode be identical with the physical state of the substance of the cathode, the cell is strictly unpolarisable. The latter condition is not realised when a strong current is produced.

317. *Voltaic Batteries.*—Various cells may be joined together to form a voltaic *battery.* When the positive element of one cell is joined to the negative element of the next, the cells are said to be coupled in *series*, or to be coupled for electromotive force, for, in this case, the electromotive force of the battery is the sum of the electromotive forces of the several cells. *Volta's pile* is essentially an arrangement of cells in series. A copper disc is laid upon a disc of blotting-paper or flannel, moistened with brine, which is placed on a zinc disc. Another zinc disc is placed upon the copper disc, and so on, the series ending with a copper disc. The single cell in Zamboni's *dry pile* is a disc of paper covered with zinc foil on one side, and with manganese dioxide on the other. The slight amount of moisture contained in the paper acts as the liquid of the cell.

When all the positive elements are joined together, and all the negative elements are joined together, the cells are said to be coupled for *quantity*, for the whole battery acts as one large cell of the same kind, and produces a powerful current under a relatively small electromotive force.

The best arrangement of similar cells to produce the maximum current depends upon the external resistance as well as the resistance of each cell. Let R be the external resistance, while r is the resistance of each cell and e is its electromotive force. Let there be N cells, and let there be n_1 rows of these, each containing n_2 cells in series, so that $N=n_1n_2$. Let this arrangement give a certain current; and suppose that, if n_2 be increased by unity, all the N cells being still used, the same current is again given. Either arrangement is then as good as any which can be obtained. The total electromotive force in the first case is n_2E, and in the second it is $(n_2+1)E$. The total resistance in the first case is $n_2{}^2r/N+R$, and in the second it is $(n_2+1)^2r/N+R$. Hence, by Ohm's law, we get

$$n_2(n_2+1)r=NR.$$

This means that the external resistance exceeds the internal resistance by one n_1th part of the resistance of a single cell. In order that the whole number of cells may be employed in both cases, we must also have $n_1=p(n_2+1)$, where p is a whole number. That is, r must be a whole multiple of R. Other cases may be similarly worked out.

If n_2 be large, the condition is practically $n_2r/n_1=R$; that is, the internal and external resistances are equal.

The best arrangement for greatest efficiency is the arrangement in multiple arc; for the internal resistance has then its least value, and the least amount of energy is wasted in overcoming it.

318. *Secondary Cells.—Grove's Gas Battery* was, with the exception of Ritter's (1803), the earliest form of secondary cell. It consists essentially of two glass tubes (A, B, Fig. 169), which are fitted into the two necks of an ordinary Woulfe's bottle. These tubes are open at their lower ends, but are closed at their upper ends. Platinum wires are inserted through the glass at the closed ends, and are welded to strips of platinum foil, which extend throughout nearly the whole length of the tubes. The tubes and the bottle are filled with (say) dilute solution of sulphuric acid.

If a current be sent round the circuit from A to B through the liquid, oxygen will be evolved in A and hydrogen will be evolved in B; and a reverse electromotive force will be produced. If the direct current be stopped, and if the wires let into A and B be joined, this reverse force will produce a reverse current setting from

B to A through the liquid. The liquid will be decomposed, oxygen tending to appear in B and hydrogen tending to appear in A. The gases are not really set free; they combine with the gases already in the tubes to form water; and this process goes on until both tubes are again filled with the liquid solution.

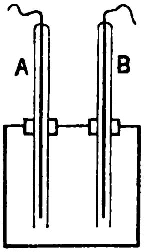


FIG. 169.

[It is interesting to note that the arrangement may be used as a primary battery. Instead of evolving the gases by a direct current, we may introduce them independently into the tubes, and the current will proceed from B to A through the liquid as formerly.

The current will flow even if hydrogen be introduced into B while A is left full of liquid. The hydrogen in B unites with the oxygen which is produced by the electrolytic action of the current, and hydrogen is evolved in A. The sum of the electromotive forces in the circuit soon becomes zero, since there is hydrogen in both tubes, and the current stops.

If A be filled with ordinary oxygen while B is left full of liquid, very little effect will be observed; and this shows that electrolytically evolved oxygen differs considerably in its properties from ordinary oxygen. If A contains ozone, the action will proceed as before.]

Planté's secondary cell consists of two large sheets of lead, separated by a sheet of guttapercha, and rolled up into a spiral form. These sheets are placed in a vessel which contains dilute sulphuric acid, and an electric current is passed through the cell from one sheet of lead to the other. The oxygen which is evolved at the one sheet forms a layer of peroxide of lead on its surface. The direct current being stopped, and the terminals of the lead sheets being joined, a reverse current flows through the cell. The hydrogen which is now evolved partly reduces the peroxide of lead, and the oxygen unites with the lead of the other electrode to form peroxide;

and, when both electrodes become similar, the current stops. A current is now passed, from an external source, through the cell in the same direction as that in which the reverse current flowed. When the gases bubble off freely at the electrodes, this current is stopped, and the polarization current is allowed to flow; and so on alternately for a number of times. This process, which is said to *form* the cell, gradually changes the lead into a spongy condition, and thus greatly increases the charge which the cell can contain. When the process of formation is complete, the cell is always charged by a current in one direction only.

Fauré's cell is essentially similar to Planté's, but the plates of lead are at first covered with a coating of a lower oxide of lead. When the direct current is passed, the oxide on the one plate is changed to the peroxide, while that on the other is reduced to the metallic condition. This avoids the tedious process of forming the cell by alternate currents in opposite directions.

The modern *accumulator* is constructed on essentially the same principles as those on which the Faure cell is constructed, though various improvements are introduced. Its electromotive force exceeds that of a Bunsen cell, while its resistance is extremely small, so that it is capable of producing a very powerful current. It is an accumulator of *energy*, not an accumulator of electricity.

319. *Transformations of Electric Energy in Conducting Circuits.*—If the potential V of a conductor be kept constant while its charge alters by the amount Q, the electric energy changes by the amount VQ. Hence, if the quantity Q of electricity flows from a part of a conductor where the potential is V to a part where the potential is V′, the energy expended in the production of this flow is $(V-V')Q$, that is, it is EQ, where E is the electromotive force which acts between the given parts. Hence, if a current of intensity C is maintained between these parts, the rate at which energy is expended in the process is EC.

This energy, which is associated with the current, will be transformed into heat in the circuit, if no other transformations take place. If a reverse electromotive force E_0 acts in the circuit, work is expended at the rate E_0C in making the current flow in a direction opposite to that in which the reverse force acts. And this energy is transformed into heat at the place where the reverse force acts. (The Peltier evolution of heat at a thermo-electric junction is a case in point.)

If no part of the direct electromotive force acts against a reverse force, and if R is the resistance of the circuit, Ohm's law gives

E = CR, whence we see that the rate at which heat is developed in the circuit is, in electrical units,

$$RC^2;$$

that is to say, *the rate at which heat is developed in a circuit is directly proportional to the resistance of the circuit and to the square of the intensity of the current which flows through it.* This is known as Joule's Law.

Part of this heat is produced in the cell which is used for the production of the current, but by making the external resistance very large in comparison with the internal resistance of the cell, this portion may be made as small a fraction of the total amount as we please.

This principle is utilized in the process of electric lighting. The carbon filament of an 'incandescent' lamp is of relatively large resistance. Almost all the heat into which the electric energy is transformed is developed in the carbon, which becomes 'white-hot.' In the 'arc' lamp the chief part of the resistance is at the air-gap between the carbon poles, and the heat which is developed raises the air to so high a temperature that it becomes intensely luminous. [Air at ordinary temperatures is an insulator, but hot air admits of the passage of electricity through it with comparative ease. The heating of the air is effected by allowing the carbons to touch each other, so that the current flows through them and makes them red-hot near the point of contact. When this results, the carbons may be separated to a slight extent, and the current will continue to flow.]

The energy of an electric current may be directly transformed into mechanical work by means of an electro-motor (Chap. XXIX.).

320. *Measurement of Electromotive Force, Current Strength, and Resistance.*—The electromotive force which acts between two parts of a conductor may be determined directly by means of the electrometer (§ 298), for the problem is merely one of the determination of difference of potential.

The strength of the current which flows in a circuit may be found directly by placing an electrolytic cell (or voltameter) in the circuit. The quantity of electricity which passes through the cell per unit of time is, by Faraday's law, directly proportional to the amount of chemical decomposition which takes place per unit of time. [The arrangement which was described as Grove's gas battery, in § 318, constitutes a voltameter, and the amount of oxygen or of hydrogen which is evolved per unit of time in either inverted tube can be measured directly. If Q be this quantity, while q is the electrochemical equivalent of the substance, the strength of the current is Q/q.]

In practice the *galvanometer* is more frequently used for the measurement of the strength of a current. It consists essentially (Fig. 170) of a coil of wire within which a magnet is freely suspended. The magnet lies normally in the plane of the coil, but, when a current flows round the coil, the magnet tends to place its length at right angles to the plane of the coil. This tendency is opposed by the action of a constant external magnetic force, and the tangent of the angle of deflection of the magnet is proportional to the strength of the current (Chap. XXVIII.).

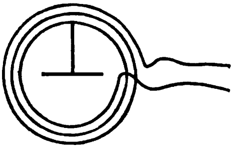


FIG. 170.

If the galvanometer be used as a shunt between two points of a conducting circuit, the total current is divided between the galvanometer and the part of the circuit included between the two points, the strength of the current in the galvanometer bearing to that of the current in the part of the conductor the ratio of the resistance of the latter to the resistance of the former. Hence, if the resistance of the galvanometer be very great, its use as a shunt will make very little change in the current in the conductor, and the galvanometer will practically indicate the difference of potential between the two given points of the conductor. When used in this way, the galvanometer is termed a *high resistance* or *potential* galvanometer.

If the magnet of the galvanometer be replaced by a movable coil of wire through which the current passing through the galvanometer itself also flows, the indications are proportional to the *square* of the strength of the current. Instruments constructed on this principle are often called *electro-dynamometers.*

A galvanometer which is graduated so as to directly indicate the strength of the current in amperes is called an *ampere-meter* or *ammeter;* and any instrument, whether constructed on the above principle or not, which indicates difference of potential in volts, is called a *voltmeter.* In Cardew's voltmeter the current traverses a long thin platinum wire, which is stretched by a constant tension, and the difference of potential is measured by means of the change of length of the wire which results from the heating effect of the current (§ 319).

The resistance of a wire may be determined in terms of any given unit by means of the arrangement known as *Wheatstone's Bridge.* Let four conductors, the resistances of which are r_1, r_2, r_3, and r_4, be arranged as in Fig. 171, and let a battery b be placed between the points A and B, so that currents flow (say) in the way which is indicated by the arrows. The potential at the point C is intermediate between those of A and B, since the current flows, through C, from A to B. And, by Ohm's law, since the same current flows along AC and CB, the potential at C must divide the difference of potential between A and B in the ratio of r_1 to r_2. Similarly the potential at D will divide that difference in the ratio of r_3 to r_4. Hence, if C and D are at the same potential (which may be determined by the fact that no current will then flow through a galvanometer which is placed between C and D), the resistances must be connected by the relation

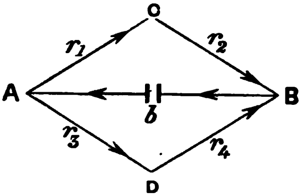


FIG. 171.

$$\frac{r_1}{r_2}=\frac{r_3}{r_4}.$$

And consequently, if we know the ratio r_3/r_4, and also the absolute value of r_2, we can calculate the value of r_1.

The numbers in the following table give approximate values of resistances, per cubic centimetre, in ohms.

Guttapercha (24° C.)	$45\ (10)^{13}$
Glass (20° C.)	$9\ (10)^{13}$
Sulphuric Acid (30 % sol. at 18° C.)	$140\ (10^{-2})$
Mercury	$943\ (10)^{-7}$
Manganin	$475\ (10)^{-7}$
Platinoid	$325\ (10)^{-7}$

German Silver	$208\ (10)^{-7}$
Iron	$964\ (10)^{-8}$
Platinum	$898\ (10)^{-8}$
Aluminium (annealed)	$289\ (10)^{-8}$
Silver (hard)	$162\ (10)^{-8}$
Copper (hard)	$160\ (10)^{-8}$
Copper (annealed)	$157\ (10)^{-8}$
Silver (annealed)	$149\ (10)^{-8}$

The values, in the cases of the metals and alloys, are those obtained at 0° C. The increase of resistance, per degree Centigrade, is about 0·4 per cent. in the case of metals. In the case of German silver the coefficient is about 10 times smaller; and, in the case of platinoid, it is about 40 times smaller. Dewar and Fleming have shown, by measurements at the lowest attainable temperatures, that the resistances of pure metals diminish, with fall of temperature, as though they would vanish at the absolute zero.

If we know the value of any two of the three quantities—electromotive force, current-strength, and resistance—we can calculate that of the third by means of Ohm's law ($E=CR$). Or, if H be the heat which is developed in a conductor by transformation of electric energy, while J is the mechanical equivalent of heat, we can calculate the values of any two of the three quantities, when we know that of the third, by means of Joule's law,

$$EC=RC^2=JH.$$

The resistance of a metallic conductor increases when its temperature is raised, but the resistance of an electrolytic conductor diminishes when its temperature increases. The law of variation being known, we can determine temperature by means of measurements of resistance. This method is very useful when the temperature is high.

A discussion of the various units in terms of which electrical quantities are measured will be given in Chap. XXIX.

The variation of resistance with temperature forms the basis of the most delicate method for the measurement of radiant energy. The *bolometer* which is used for this purpose consists essentially of an extremely sensitive and well-balanced Wheatstone's Bridge. This bridge is thrown off balance, and a current is produced through the galvanometer, when radiant heat falls on one of the resistances.

To measure the true resistance of an electrolytic conductor special methods must be employed. The best of these is one introduced by Kohlrausch. The Wheatstone Bridge arrangement is employed, and alternating currents are used, so as to prevent the development of polarisation which ensues when the current is continuous. If an ordinary galvanometer is employed, the alternating current must be made continuous, before it enters the galvanometer, by means of a special commutator. Otherwise, an electro-dynamometer, or a telephone, is used instead of an ordinary galvanometer.

321. *Conductance expressed as a Velocity.*—Suppose that we have a sphere which can expand and contract, and let E be its potential, while a is its radius. Its charge is then $Q=Ea$. If it be connected to the ground by a wire of conductance K, the charge will diminish, but the potential will not fall if the radius diminishes at a suitable rate. Since the ground is supposed to be at zero potential, E is the electromotive force acting along the wire. By Ohm's law the strength of the current is EK. But the strength of the current is also the rate at which the charge diminishes, and this is the product of E into the rate at which the radius diminishes. Hence the conductance is numerically equal to the inward velocity of any point on the charged sphere under these conditions.

322. *Currents of Varying Strength.*—The laws of conduction which have been considered in this chapter apply only to the case of steady currents. The phenomena which are observed when the strength of the current varies are more complicated than those here described. Consideration of them will be made in Chap. XXIX.

CHAPTER XXVII.

ELECTRIC CURRENTS IN GASES.

323. *Molecular or Atomic Charge.*—In discussing the question of the conduction of electricity through solids in the preceding chapter, no inquiry was made regarding the mechanism of such conduction. Neither was this inquiry made when the phenomena of conduction through electrolytes were dealt with. But the laws of electrolytic conduction pointed distinctly to an intimate connection between electricity and matter. And, as we do not yet know what electricity is or what matter is, further investigation of the connection between them is the most promising path by which we may either seek to arrive at some understanding of the nature of each or attempt to obtain foundation for a suitable hypothesis.

Now, the gaseous condition of matter is the condition with regard to which we have greatest knowledge, and so the investigation of the laws of the passage of electricity through gases is an investigation of the greatest importance.

In so far as the passage is strictly passage through a gas, and not passage by means of solid particles suspended in the gas and giving rise to convective discharge (§ 308), the medium of that passage may be the molecules, or the atoms, or both.

Now, the charge of an electrified insulated body cannot be retained when the body is surrounded by dusty air. On the other hand, if the air be free of dust, the charge can be retained when the temperature is low. But, at a low temperature, there are very few free atoms present. We therefore arrive at the conclusion that, at low temperatures, the leakage of a charge in dusty air is due to the dust—the molecules bearing no part in the process. Though we cannot logically arrive, from this datum, at the conclusion that a molecule cannot carry a charge in any circumstances, many facts indicate that such is the case—for example, the fact that the molecules which evaporate from an electrified liquid do not carry a charge with them.

When the temperature is high, a charge cannot be retained even in dust-free air. This might be due to the air—originally dust-free—really ceasing to be dust-free when the temperature is high. Dust might be produced by slight disintegration of the charged body. On the other hand, it might be due to dissociation of molecules into atoms, provided that free atoms can carry charges. Various facts indicate that the latter explanation is correct.

Gases may be divided into two well-marked classes, according as, when hot, they conduct electricity well or badly. It is found that a hot gas conducts readily if it is dissociated into *atoms* by means of heat, while it does not conduct readily if its molecules only dissociate into simpler molecules. Hydriodic acid gas is an example of the former kind, ammonia gas is an example of the latter kind. Conduction, also, does not occur readily in a heated gas if there is no dissociation of molecules.

The slight conduction which occurs in the two latter cases might be due to dust produced by disintegration of the electrodes, or to atoms produced by chemical action at the electrodes. It will cease if the electrodes are kept cold, even though the gas be hot.

324. *Effect of Violet Light.*—(1) If violet light falls on the cathode, provided that the cathode be metallic or a phosphorescent liquid, conduction will take place even in a cold gas. The readiness with which this effect appears depends on the pressure and the nature of the gas. It is most easily produced by invisible rays beyond the violet end of the spectrum, yet it may be produced by ordinary luminous rays. Thus, the diverging leaves of a negatively charged electroscope will collapse if the flame of a Bunsen burner be held in the neighbourhood of an insulated metallic body which is connected to the electroscope by a metal wire.

In these cases it is found that a steam jet, in the neighbourhood of the cathode, will condense. Hence the action may be due to dust particles produced by disintegration of the cathode under the influence of the light (see § 176), or to chemical change occurring near the jet, for both causes lead to condensation of water vapour.

(2) Again, an insulated and uncharged metal, especially freshly polished zinc, if exposed to ultra-violet light, becomes positively charged. The more powerfully a body absorbs the light, the more readily does the action ensue, and the more electro-positive the substance is the more readily does it ensue. Phosphorescent liquids (which phosphoresce *because* of the absorption of light) are acted upon similarly.

The 'negative rays' (§ 335) from a cathode, which produce phosphorescence, also cause the production of a positive charge in an insulated uncharged metal.

In these cases the action might be due to disintegration of the metal or liquid surfaces, provided that disintegrated dust takes away a negative charge with it. In support of this view there is the well-known fact of the ready disintegration, or 'spluttering,' of a negatively charged metal.

(3) An insulated, positively charged, body will lose its charge if it be placed in the neighbourhood of an illuminated uninsulated metal. This farther supports the explanation given with regard to (2) above.

325. *Effect of Glowing Metals.*—(1) In air or oxygen, at a low pressure, negative electricity is discharged more easily from a cold metal than positive is, when a bright red platinum wire is placed in the neighbourhood. (2) In hydrogen, at a low pressure, the opposite effect occurs. (3) A white-hot iron sphere will, in air, retain neither kind of electricity; but, as it cools, it will retain negative electricity before it can retain positive electricity. (4) An insulated body when in the neighbourhood of a hot platinum wire becomes positively charged in air. (5) In hydrogen, it becomes negatively charged. (6) If a thin platinum wire be kept at a red heat for a long time, an insulated body in the neighbourhood becomes negatively charged in all gases.

With the exception of the results which are first obtained in air and oxygen, these phenomena may be explained on the assumption that the vapour of a metal has a greater attraction for negative electricity than the solid metal has. And the action described in (6) indicates that the first effects obtained in air and oxygen are merely an exception to a general rule. J. J. Thomson suggests that these effects are due to little solid particles torn away from the hot metal, which, being solid, have the same attraction for electricity as the mass of metal has, and so carry away some of the charge of the metal. Thus the action of the solid particles is contrary to that of the vapour, and will preponderate until the disintegration into solid particles largely ceases. In support of this explanation there is the fact that platinum does not 'splutter' in hydrogen gas.

326. *Electric Strength, Difference of Potential, and Spark Discharge.*—We have already seen (§ 294) that, when the difference of potential of the electrodes exceeds a certain value, depending on various circumstances, the insulating power of the medium breaks down and a spark discharge occurs. Maxwell defined the *electric strength* of the medium as the maximum electromotive intensity (electromotive force per unit length taken in the direction of the resultant force) which the medium can sustain without the occurrence of a discharge. Various facts indicate that there is no fixed property of a medium corresponding to electric strength so defined.

The difference of potential seems to be a more fundamental condition than electromotive intensity. The electric strength depends on the material of which the electrodes are made; discharge takes place more easily with aluminium electrodes than with electrodes of other metals. It depends on the size and shape of the electrodes; in the case of two equal spheres it increases with the curvature. It depends upon the distance between the electrodes; there is a certain distance at which, other things being equal, the breaking stress is least. It depends upon the state of the surface of the electrode, and the condition of the electric field as regards uniformity. Small rapid changes in the potential of an electrode cause more ready discharge.

327. *Spark Length and Potential Difference in a practically Uniform Field.*—When the spark length is not too small, the potential difference increases, with increase of spark length, at a practically constant rate such that the excess of the potential difference over a certain value is proportional to the spark length. As the spark length becomes smaller, the potential difference diminishes at first at a greater and greater rate, then at a smaller and smaller rate, reaches a minimum, and finally increases as the spark length is still further lessened. The latter peculiarities can only be made evident under diminished pressure. At atmospheric pressure the spark lengths at which they presumably occur are excessively small.

328. *Spark Length and Potential Difference in a Non-uniform Field.*—When the field is not uniform—a condition which may be realized by using small spherical electrodes—the general nature of the relation between spark length and potential difference is the same as that in a uniform field. If spheres of a given diameter be replaced by smaller spheres, there is a definite spark length at which the potential difference is not altered by the substitution. At smaller spark lengths than this, greater potential difference is required with the smaller spheres; at larger spark lengths less potential difference is required. If still smaller spheres were used, the same rule would hold, but the definite spark length at which the substitution would cause no change of potential difference would be less than in the former case. Hence, for any given spark length, there is a critical size of sphere which corresponds to the maximum potential difference.

If the spark length be not too small, the potential difference which is required to produce a spark of given length is smaller in a non-uniform field than it is in a uniform field. But, with small enough spark lengths, less potential difference is needed in a uniform field than in a non-uniform field.

329. *Maximum Electromotive Intensity and Spark Length.*—The

electromotive intensity at which sparking takes place is very large when the spark length is small. It diminishes as the spark length increases—at first very rapidly, then more slowly—until it reaches a minimum value, after which it goes on increasing as the spark length increases.

At a given spark length, increased non-uniformity of the field causes increase of the maximum electromotive intensity; and the change which is produced by irregularity of the field is much greater than that which is produced in the potential difference by the same irregularity of field. This is one of the facts which indicate that potential difference plays a more important part than does electromotive intensity in spark discharge.

When the field is uniform, the potential falls uniformly from a certain value at the anode to (say) zero at the cathode. When the spark passes, the potential of the anode may be maintained at the value which it had when the insulation first broke down, while the potential of the cathode assumes a certain value. The potential gradient is now less than before, but is still uniform.

On the other hand, when the field is not uniform, the potential gradient before discharge will be practically uniform in the near neighbourhood of the anode but will be smaller than it was in the uniform field. At greater distances from the anode the gradient becomes greater and greater, reaching, at the cathode, a value greater than that which it had in the uniform field. It is, therefore, possible that, in this non-uniform field, the potential may, at some point between the electrodes, reach a value equal to that which exists at that point during discharge, even although the potential of the anode has not reached the value required for discharge in a uniform field. Thus we can understand how the total potential difference, at a given spark length, requisite for discharge in a uniform field may be greater than that requisite in a non-uniform field; and also how the breaking stress may be less in the former case than in the latter.

330. *Spark Potential and Gaseous Pressure.*—If the spark length be kept constant, while the pressure of the gas through which the spark passes is diminished, the potential difference necessary for the passage of the spark first diminishes and then increases. The critical pressure at which the potential difference has its minimum value depends on the nature of the gas and on the shape and size of the electrodes and of the vessel in which the gas is contained. These conditions being fixed, it varies very markedly when the spark length is altered.

At pressures considerably over the critical pressure, the increase of potential difference is nearly proportional to the increase of pressure. The ratio of the former to the latter increment really decreases slowly as the pressure rises.

If the product of density and spark length is kept constant, the potential difference requisite for sparking remains nearly constant throughout a great range of pressures. This means that, when Boyle's Law practically holds, the product of the spark length and the pressure is nearly constant. And this in turn implies that *the ratio of the spark length to the mean free path is nearly constant when the potential difference is constant*, a result of the greatest theoretical importance (§ 341). It is found by experiment that the potential difference, when it is varied, is nearly proportional to this ratio.

Although the critical pressure corresponding to minimum potential difference decreases very greatly as the spark length increases, the value of the minimum potential difference varies very little, a result which emphasises the importance of potential difference, rather than electromotive intensity, in the production of sparks.

When the spark length is small, the potential difference does not vary much throughout a large range of pressures. The reason for this will appear in § 336.

The existence of the critical pressure does not essentially depend on the presence of electrodes. It is evident when the discharge is made without electrodes (§ 331), although the pressure has, in that case, to be reduced, in order to reach the minimum potential difference, to a lower value than that which is necessary with electrodes.

When electrodes are used, increase of the diameter of the tube containing the gas lowers the critical pressure.

331. *The Electrodeless Discharge.*—Before proceeding to consider in greater detail the phenomena which accompany the passage of electricity through a vacuum tube provided with electrodes, it will be well to consider, because of the greater simplicity of conditions, the leading phenomena of the electrodeless discharge.

We shall find subsequently (§ 363) that, when a current of varying strength flows through a conductor, a current, also of varying strength in general, is induced in a neighbouring conductor. This principle is employed in the production of the electrodeless discharge.

If a glass tube (called a *vacuum tube*), containing a gas under diminished pressure, be placed within a coil of wire which is included in the secondary circuit of a powerful induction coil (§ 365), a ring of light will be formed inside the tube, in the region nearest the surrounding coil of wire, when the induction coil is worked.

This light is caused by the electrodeless discharge which results from the electromotive force which is induced in the conducting-gas by the varying currents in the surrounding coil. In the case of air, the luminosity has a maximum value at a pressure of about $\frac{1}{200}$ of a millimetre of mercury. This is a pressure about one hundred times less than that which gives greatest luminosity when electrodes are used.

The existence of the critical pressure may be very readily shown by means of this apparatus. If the vacuum tube contain mercury and mercury vapour, the pressure of the mercury vapour may be varied by the application of heat. It is found that the discharge will only pass when the temperature ranges from about 70° C. to 160° C.

332. *Molecular Conductivity of Gases.*—If the vacuum tube which is placed inside the coil of wire be so constructed that another vacuum tube can be placed within it, the electrodeless discharge will take place in the inner tube if the outer tube contain air at atmospheric pressure. But the pressure of the gases in the two tubes may be so arranged that the discharge will take place in the outer tube rather than in the inner tube. Let the pressure of the gas in the outer tube be such that the discharge just ceases in the inner tube. Now fill the outer tube with a solution of sulphuric acid (say) of such strength that the discharge just fails to pass in the inner tube. Under these conditions, the acid solution and the gas respectively have the same conductance. Since the pressure of the gas is known, the (relative) number of molecules which it contains per unit volume is known; and, since the strength of the acid solution is known, the number of acid molecules per unit volume is also known. Thus we can compare the conductivities per molecule of the gas and the solution. It is found that, although the conductivity of a gas is much less than that of a metal, the molecular conductivity of a gas is excessively large in comparison with that of a metal.

333. *Passage of Electricity between Dissimilar Substances.*—We can readily obtain evidence that there is considerable difficulty in effecting the passage of electricity between a metal and a gas. If a metallic obstruction be placed in the path of the discharge in a vacuum tube—say, the electrodeless discharge—the electricity will take a longer path rather than pass through the metal. It has been shown also, by Liveing and Dewar, that, in the light of the discharge in a vacuum tube the gas in which holds metallic dust in suspension, no trace of the spectrum of the metal is observed. This result is at least in accordance with the former one.

It is therefore to be expected that the presence of metallic electrodes will render the discharge more difficult, and will make necessary the expenditure of a considerable amount of energy in order to make the electricity pass from the metal to the gas, or from the gas to the metal.

It is found also that, in a mixture of two gases, the spectrum of one may be obtained alone—that of the other not appearing—when the discharge passes. Crookes has also found that, in a mixture of two gases, groups of striæ (§ 334) may be observed, the bright part of one giving the spectrum of one gas while the bright part of a different stria gives the spectrum of another gas. These facts seem to indicate that there is also difficulty in the passage of electricity from one gas to another.

334. *Discharge with Electrodes.*—In Fig. 172 (from a photograph) an indication is given of the luminous phenomena which are observed when the discharge passes between electrodes in a vacuum tube. The negative electrode is represented as a straight wire entering the tube at the left-hand side of the diagram. Over its surface there is a bright glow, more or less in patches, which can cast a shadow towards the electrode. Outside this there is a dark space which is known as the *first dark space*, or as *Crookes' space.* The length of this space increases with the mean free path of the molecules, but not in proportion to it. There is next a bright space, called the *negative glow*, the length of which is variable. According to Crookes, this is a region in which occur collisions of particles shot off from the negative electrode, and the first dark space is a region in which no such collisions take place. In support of this view, there is the fact that the shape of the bounding surface between these regions may be obtained by finding the locus of lines of constant length drawn perpendicularly from the surface of the negative electrode. Also any substance placed in the first dark space will stop the negative glow in the region behind it, and, if it be phosphorescent, will itself be made to glow. If the first dark space extends to the walls of the tube, the glass is made to phosphoresce —with blue light if it be lead glass, with green light if it be German glass; and a sharp shadow of any substance placed in the space will be thrown on the walls.

The negative glow is generally succeeded by the *second dark space*, or *Faraday space*, which is not of fixed length, and is sometimes absent.

Then comes a succession of luminous striæ which occupy the remaining space up to the positive electrode. A few of these are shown in the diagram. The diagram must not be supposed to give more than a mere indication of the succession of brightness and darkness.

This succession of striæ constitutes the *positive column*. Each stria is in general distinctly concave towards the positive electrode, a fact which the figure does not indicate well. The distance between the bright parts of successive striæ depends on the pressure of the gas and the diameter of the tube. It increases when the diameter increases, provided that the striæ reach out to the walls of the tube; and it increases as the pressure diminishes. The ratio of the values of this distance at two given pressures is quite independent of the size of the tube.

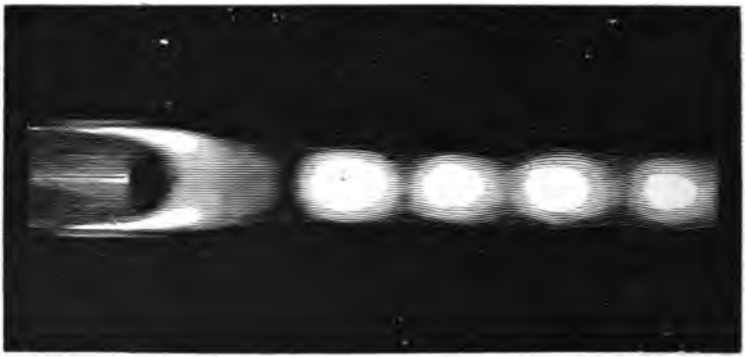


FIG. 172.

The striæ move rapidly backwards and forwards in the tube to some extent, and sometimes the motion is so great that the striæ apparently disappear. In this case their existence may be made evident by the use of a rotating mirror. There is a certain strength of current which gives maximum steadiness.

If the negative electrode be moved forward in the tube, the whole system of bright and dark parts move forward as if they were rigidly joined to it. If the positive electrode be moved forward, successive striæ disappear as it passes them. If the anode be placed in a side branch of the tube, the positive column may be confined to the side branch, but the first dark space and the negative glow is always confined to the other part of the tube, and will extend right past the opening of the side branch. In fact, the phenomena occurring at the cathode are local, and will disappear if the cathode becomes incandescent. It appears, therefore, that the discharge is conveyed by the positive column.

J. J. Thomson has found, by means of a rotating mirror and a long discharge tube, that the luminosity starts from the anode and travels to the cathode at a speed about equal to one half of that of light, the speed being independent of the shape or nature of the electrode.

335. *The Negative Rays.*—We have seen that there are appearances in the neighbourhood of the negative electrode which are in agreement with the supposition that they are due to particles shot off from the electrode and following rectilinear paths like rays of light. The term 'negative rays' is therefore usually employed in the discussion of the subject.

Crookes' supposition that the rays are the paths of negatively-charged particles is further supported by the fact that two pencils of such rays exert mutual repulsion. On the other hand, a charged body does not deflect the rays; but, as the gas is a good conductor, and as a good conductor screens electrostatic action (§ 291), in all probability the action is prevented from reaching the rays.

The charged particles are not simply little parts torn off from the metallic electrode, for the rays can be produced if the electrodes are placed on the outside of the vacuum tube, so that the discharge is due to the inductive action of the electrification of the electrodes. They are probably atoms, as Schuster has shown. If a vacuum tube be arranged so that both the electrodeless discharge and the discharge with electrodes can take place in it, and if the conditions be so disposed that the electrodeless discharge is just unable to pass of itself, the electrodeless discharge will at once occur if the other discharge be started. Further, if the discharge with electrodes takes place partly in a side branch of the tube, and if the negative electrode be so situated in that branch that the rays cannot reach the region in which the other discharge is made, the latter will not occur. It seems therefore that the negative rays provide that supply of dissociated atoms which is essential to discharge through gases.

A metallic body receives a positive charge from these rays; and this fact at first sight seems to directly disprove the assumption that they are caused by negatively-charged particles. But the fact of the great difficulty which attends the transference of electricity from a gas to a metal, leads rather to the idea that such transference does not occur in this case, and that the positive charge of the metal is due to the same cause as the positive charge given to a metal under the action of ultra-violet light. This view is supported by the fact that the negative rays and ultra-violet light both give rise to phosphorescence when they fall upon a phosphorescent body.

If the negative rays actually correspond to the paths of particles, vanes, having very frictionless bearings, might be expected to be set in rotation when properly placed in a pencil of rays. This result has been observed; but, on the other hand, the rotation may be due to the development of heat in the vanes, as in the radiometer.

336. *Variation of Potential from Anode to Cathode.*—It is probable that in most cases the current in a vacuum tube is intermittent, the total difference of potential gradually rising until it is sufficient to produce discharge, when it falls, and then again rises until another discharge takes place, and so on. The difference of potential between any two points may be found by connecting an electrometer to platinum electrodes let into the vacuum tube at these points; but if the discharge be intermittent, the instrument only indicates an average difference.

If the pressure of the gas is of such a value that the negative glow does not extend to the walls of the tube, the total difference of potential from anode to cathode remains nearly constant, though the strength of the current is varied greatly. After the glow reaches the walls, the total difference increases markedly and continuously as the pressure is further diminished. This is observed even in the electrodeless discharge; but, in the discharge with electrodes, the variation is in large part confined to the neighbourhood of the cathode. In the positive column, the potential difference diminishes as the pressure diminishes, and tends ultimately to become constant. In the Faraday space the potential difference is abnormally small.

Hittorf showed that the fall of potential in the neighbourhood of the cathode does not depend upon the strength of the current so long as the negative glow does not extend over the whole of the cathode; and Warburg showed that, under the same condition, it does not depend upon the pressure.

When the negative glow extends over the whole of the cathode, the temperature of the gas at the cathode increases when the current is increased, a condition which agrees with the fact of the increase of potential difference and the corresponding expenditure of energy. The heat which would be developed in accordance with Joule's Law (§ 319) is far too small to account for the elevation of temperature.

All luminosity disappears if the cathode, or even the anode, is made white hot, and a white-hot platinum wire placed in the circuit will destroy the luminosity in its neighbourhood; and it is found that, when the cathode is white hot, a small potential difference will maintain the current, although a spark at high potential might be required to start it. Thus the luminosity is associated with high potential difference or great expenditure of energy.

In all the above cases the potential difference requisite to *maintain* the current is the quantity under consideration.

If the electrodes be placed so near together that the dark space at the negative electrode extends beyond the positive electrode, the discharge proceeds from the back part of the anode, along the surface of the glass, to the negative glow at the back of the cathode. And if, under this condition, there are two possible paths, one long and one short, by means of which the discharge can go, a very long path will be followed rather than the direct one.

In order to account for the variations of potential along the tube, it is essential that there must be a distribution of free electricity in the tube when the discharge is going on.

337. *Variations of Pressure during Discharge.*—When the discharge takes place in a vacuum tube, there is a sudden increase of pressure, which may amount to about 30 per cent. of the whole pressure, and so is not due to the expansion consequent on the development of heat. Neither is it due to the increase of pressure which follows decomposition of molecules into atoms, for the amount of decomposition is not nearly sufficient. It may be due to the sudden propulsion of electrified particles.

If the vacuum tube be placed between the plates of a condenser, the above effect is observed when the condenser is either charged or discharged, so that it does not essentially depend on the passage of a spark.

338. *The Arc Discharge.*—If two carbon electrodes be placed in contact, and a sufficiently strong current be sent through them, intense incandescence will be produced. If the electrodes be then separated slightly, the discharge will pass in the form of an arc stretching from one electrode to the other through the air, although the difference of potential was too small to originate the discharge with the given air-gap. The distribution of potential is represented by the formula $V = a + bl$, where l is the length of the air-gap. The value of the quantity a depends on the nature of the electrodes, being smaller the more volatile they are. When carbon electrodes are used, there is much more disintegration at the anode than there is at the cathode. The anode soon becomes hollowed out.

In all probability the disintegration occurs by thermal processes as well as by electrical processes, so that no definite relation can be expected to hold between the loss in weight of the electrode and the amount of electricity which has passed. An experiment by Grove, made with a zinc electrode so massive that the thermal effect must have been small, seemed to show that the amount of zinc which was lost was chemically equivalent to the amount of oxygen which was simultaneously evolved in a voltameter placed in the circuit. On the other hand, Hering did not get this result when he used a massive silver electrode. Fleming supposed that the arc discharge was carried by disintegrated particles torn from the anode, but this does not agree with the known fact that the ordinary discharge with striæ can be made to pass continuously into the arc discharge.

In the arc discharge the temperature at the anode is greater than the temperature at the cathode. In a vacuum tube at a high pressure this is also the case, though at a low pressure the reverse is true.

339. *The Discharge in a Magnetic Field.*—In Chap. XXIX. it will appear that a flexible conductor carrying a current is in general made to change its position when a magnetic field (§§ 342, 343) is produced in its neighbourhood. It is also found that the discharge in a vacuum tube in general suffers displacement when it occurs in a magnetic field, and the displacement is precisely that which would take place if it consisted of a series of flexible conductors. It must be noticed that, in a conductor, the lines of flow are not displaced—the conductor is displaced as a whole. Now, in the vacuum tube, the whole body of gas filling the tube forms the conductor; and, as its glass boundary is fixed, we might have expected that no displacement of the discharge could occur. But the motion of a body which has an electric charge constitutes an electric current. Hence, if the flow of electricity in a gas is to obey the known laws of electrical flow, we must admit that the discharge is produced by the motion of electrified particles or by transference through chains of polarised particles. Confirmation of this view was obtained by Feddersen, who showed that a rectilinear discharge between electrodes in air could be blown into a curved path by a transverse blast of air. This was shown by means of the rotating mirror.

The striæ are displaced as if they carried a current of positive electricity from anode to cathode, and each stria acts as if it corresponded to a separate discharge.

The negative rays are deflected as if they carried a negative charge from the cathode. If the cathode be surrounded by a glass tube which is open only in the direction of its length, the pencil of negative rays which passes out of the tube is twisted into spirals or rings, according to the relative direction of the rays and the magnetic field. J. J. Thomson has shown that these are precisely the paths into which negatively-charged particles in rapid motion would be deflected under the same conditions.

The negative glow is moved over the surface of the electrode in precisely the same way as that in which the free ends of wires carrying a positive current to the electrode would be moved. The surfaces of the negative glow arrange themselves so as not to cut lines of magnetic force (§ 342). And, when the glow is driven from a region which it ordinarily occupies, the striæ move so as to occupy the region which it has left.

The passage of the discharge in a tube is aided when the direction of the field is along the tube, but is rendered more difficult when the direction of the field is perpendicular to the length of the tube. Under suitable conditions, in the latter case, the discharge may be entirely stopped. In explanation of this, J. J. Thomson suggests that under the action of the field, in the latter case, the moving electrified atoms which are requisite for transference of electricity are driven out of the line of discharge, while, in the former case, they are kept in that line.

340. *Chemical Effects.*—Various chemical effects occur under the action of the electric discharge. These may be indirect results of the development of heat, even although the average temperature be small. On the other hand, they may be the direct results, or the essential conditions of the discharge.

If a glass tube coated with tin foil on its inner surface be placed inside another glass tube which is coated with tin foil on its outer surface, and if the coatings be joined to the terminals of an induction coil in action, a succession of silent discharges will pass, in the space between the tubes, from the surface of one tube to the surface of the other. If oxygen be fed in at one end of the interspace, ozone will pass out at the other. So also water vapour may be decomposed into hydrogen and oxygen, ammonia gas may be decomposed into nitrogen and hydrogen, and so on.

The nature of the electrodes has, as we have seen, an effect upon the readiness with which the discharge occurs. Small impurities make a great change. In the cases of aluminium and magnesium the discharge will take place under electromotive forces not much greater than half of those which are ordinarily required; and these materials are readily oxidized.

341. *General Conclusions.*—In the case of the striated discharge, it seems likely that each stria corresponds to a separate discharge. The electrification of the anode is first discharged forwards through a portion of the gas, the potential at the front of this part rises until a further forward discharge occurs, and so on. In each stria molecules may be arranged in Grotthus chains (§ 313), and the length of these chains is that multiple of the mean free path alluded to in § 330.

J. J. Thomson has shown that this assumption affords a simple explanation of the leading characteristics of the potential differences which are observed during the discharge.

The most plausible explanation of the various phenomena which we have been considering rests upon the assumption that dust-free gases can only conduct if dissociated atoms be present. In other words, gaseous conduction proceeds only by gaseous decomposition, or is a process of *electrolysis*. And it may be that metallic conduction is a process of the same kind, though the accompanying conditions render the proof of the fact excessively difficult or impossible.

CHAPTER XXVIII.

ELECTRO-MAGNETISM. MAGNETISM.

342. *Magnetic Effects of Electric Currents.*—Let us suppose that a strong, steady current of electricity flows upwards through a straight vertical wire; and, let us also suppose that another steady current flows round a closed circular circuit, which is placed in a vertical plane in the neighbourhood of the straight wire, and is free to rotate about its vertical diameter. Under these conditions, the circular circuit will only be in equilibrium when its plane coincides with the plane which passes through its centre and the vertical wire. One of the two possible positions of equilibrium in this case is unstable, and the other is stable. The stable position is that in which the circular current passes upwards along the side next the vertical wire; the unstable position is that in which the current passes downwards along that side. When the circuit is placed in a vertical plane, in any position other than one of equilibrium, a definite force acts so as to set it round into the position of stability.

If the circular circuit be free to rotate about a horizontal axis, passing through the vertical wire, there is still the same single position of stable equilibrium.

The force tending to set the circuit into the stable position from any other given position has a magnitude which, the currents being constant, depends only on the distance of the circuit from the vertical wire.

Hence we see that the rectilinear circuit is surrounded by circular lines of force. And, from the conditions, we see that these lines cannot be—indeed, we may easily prove by experiment that they are not - lines of electric force, although the phenomena which we have just considered present obvious analogies to electrostatic phenomena. Thus insulated, oppositely charged, and rigidly connected bodies would act similarly in a field of electrostatic force.

We shall, therefore, for convenience of description, and for a reason which will appear subsequently (§ 349), call the lines *magnetic* lines of force. And, just as electric force is supposed to act upon electricity of two kinds, we shall suppose that the magnetic force acts upon *magnetism* of two kinds, which might be termed positive and negative respectively—positive magnetism being driven along lines of force in the positive direction, while negative magnetism is driven in the reverse direction. But it is not customary, though it would be better, to employ the terms 'positive' and 'negative' in connection with magnetism. The terms *north* and *south* are used instead. For it is found that, if the vertical current be stopped while the other current persists, the circular circuit has a position of stable equilibrium identical with that which it would have if the rectilinear vertical current still persisted and was situated to the west of the centre of the circle. Thus it appears that there are magnetic lines of force passing, on the whole, in a north and south direction in the neighbourhood of the earth's surface. It was for this reason that a *strong* current was supposed to flow in the vertical wire. If the current were not strong enough, the magnetic effects due to it would not preponderate. Magnetic effects due to electric currents are distinguished as 'electro-magnetic' effects.

That face of the circuit which turns northwards is said to exhibit north magnetism, while the other exhibits south magnetism. Therefore, if the positive direction of a magnetic line of force be defined as that in which north magnetism tends to move, we see that the positive direction of the circular lines of force which surround a rectilinear circuit carrying an electric current is related to the direction of the current according to the law of *right-handed screwing motion.* And, since any circuit may be regarded as made up of rectilinear parts, this law is of quite general applicability. Thus the circular current produces lines of force passing through the circuit in such a way that the circulation of the current and the direction of the lines are related to each other in accordance with the law. In this way the magnetic properties of the circuit are produced, for these lines may, so far as external effects are concerned, be supposed to proceed from north magnetism on the one side of the circuit and towards south magnetism on the other.

343. *Ideal Magnet. Pole, Axis, and Magnetic Moment.*—We do not require, in the above discussion, to consider exactly how the magnetism is distributed in relation to the circuit. It is only necessary that there should be two kinds, associated respectively with the two faces of the circuit, and situated at an effective distance apart, so that the system may be in general subjected to a definite turning moment when placed in a region or *field* of magnetic force.

In accordance with electrical analogies, we may extend our definition of magnetism so as to imply that the turning moment, other conditions being fixed, is proportional to the quantity of magnetism. And we shall find subsequently (§ 359) that the methods of measuring quantity of magnetism show that equal quantities of north and south magnetism are always associated.

We may assume, if we so wish, that, in the case of the circular circuit, a quantity of north magnetism is condensed at a point slightly in front of the centre of the circuit on its one side, and that an equal quantity of south magnetism is condensed at another point similarly situated behind the circuit. This arrangement constitutes an ideal *Magnet*. The points at which the quantities of magnetism are condensed are termed the *Poles* of the magnet, and the straight line drawn through the poles is termed the *Axis* of the magnet. The *Strength* of a pole is the numerical magnitude of the quantity of magnetism associated with it. The product of the strength into the distance between the poles is called the *Magnetic Moment* of the magnet. This quantity is obviously analogous to the moment of a couple (§ 53). In fact, it is numerically equal to the moment of the couple which acts upon the magnet when its axis is perpendicular to the lines of force in a magnetic field of unit strength.

344. *Magnetic Shell. Intensity of Magnetisation.*—The researches of Ampère and of Weber have enabled us to find the distribution of magnetic force in the neighbourhood of *any* conducting circuit.

It is found that a small plane closed circuit produces the same magnetic action as a small magnet, placed at some point inside the circuit with its length perpendicular to the plane, and having the direction of its axis (measured from south pole to north pole) related to the direction of the circulation of the current according to the rule of right-handed screwing motion, while its magnetic moment is equal to the area of the circuit multiplied by the strength of the current. [The word *small* means that the point at which the action is determined is very far off in comparison with the dimensions of the circuit.]

It is of no consequence in what part of the circuit the equivalent magnet is placed; and so we may assume that it is an extremely thin, uniformly magnetised magnetic shell, which fills the entire circuit. The quantity of magnetism per unit area of a magnetised surface is called the *Intensity of Magnetisation* of the surface, and so the magnetic shell is possessed of a magnetic intensity which is numerically equal to the strength of the electric current divided by the thickness of the shell. For, if I, *a*, and *t* represent respectively the magnetic intensity, the area, and the thickness of the shell, Iat is the moment of the shell, and, therefore, $ia = \mathrm{I}at$, or $i = \mathrm{I}t$, where i is the strength of the current, and It is called the strength of the shell.

Now let any finite circuit PQRS (Fig. 173) be filled with a network of infinitely small conducting meshes of any shape. Let a current i circulate in the circuit in the positive direction. We may assume that an equal current flows similarly in each of the meshes, such as *pqrs*, for the currents in each common side of two adjacent meshes exactly neutralise each other. But, as above, the magnetic effect of the current in each mesh may be supposed to be due to a magnetic shell of strength i. And hence we see that the magnetic action of the circuit PQRS at external points is equivalent to that of a magnetic shell of strength i, *and of any form*, which completely fills the circuit.

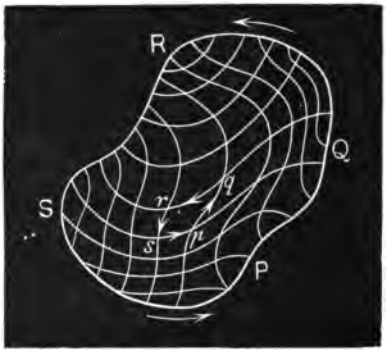


FIG. 173.

345. *The Law of Magnetic Attraction and Repulsion. Magnetic Potential.*—We can investigate, by methods to be discussed subsequently (§§ 359, 367), the law of attraction or repulsion between the quantities of magnetism which we assume to exist at the ends of magnets. The results of such measurements make it evident that the force between two quantities is directly proportional to the magnitude of each quantity, and is inversely proportional to the square of the distance by which they are separated. If we choose to regard an attractive force as negative, and a repulsive force as positive, we can symbolise this law by the equation

$$\mathrm{F} = \frac{qq'}{s^2},$$

where q, q', represent the quantities of magnetism, and s is the distance between them; for F is positive or negative according as q and q' are of like or of opposite signs.

This law is identical in form with the law of electrical attraction or repulsion, and hence all the results which we have previously deduced in the theory of electrostatics are capable of direct application in magnetostatics.

Following the electrostatic analogy, we may define the *magnetic* potential at any point, due to a magnetic pole, as the work which is expended in bringing a unit north pole—that is, a north pole of unit strength—from an infinite distance to that point. The results already deduced regarding electrostatic potential will then apply directly to our present subject.

346. *Magnetic Potential due to a Shell.*—It is easy to find a simple expression for the magnetic potential due to a shell. For if σ be an infinitesimal part of its surface, we may replace the part of the shell which corresponds to σ by a small magnet, the strength of whose pole is $i\sigma/t$. The force with which the north pole n (Fig. 174) acts on a unit north pole, placed at a distant point P, is $i\sigma/tr^2$, where r is the distance from P to n. The potential at P due to n is, therefore, $i\sigma/tr$. Similarly, the potential of s at P is $-i\sigma/tr'$, where r' is the distance from P to s. The total potential is, therefore,

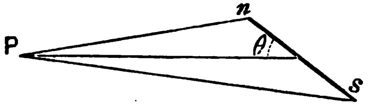


FIG. 174.

$$V=\frac{i\sigma}{t}\left(\frac{1}{r}-\frac{1}{r'}\right)=\frac{i\sigma(r'-r)}{tr^2}$$

since r' is practically equal to r. But $r'-r=t\cos\theta$, where θ is the angle between the axis of the magnet and the line joining P to its centre, and t is the length of the magnet (equal to the thickness of the shell). Hence

$$V=\frac{i\sigma\cos\theta}{r^2}.$$

Now, since the element of the surface of the shell is at right angles to the axis of the magnet, $\sigma\cos\theta$ represents the resolved part of the element normal to r, and $\sigma\cos\theta/r^2$ is the solid angle which the small surface subtends at P. We may, therefore, write

$$V=i\omega,$$

where ω represents this elementary angle. And, in order to find the total potential V′ at P, due to the whole shell, we have merely to sum all the quantities, such as V, for the whole surface. Hence

$$V' = i\Omega,$$

where Ω is the solid angle which the shell subtends at P—that is, the potential at any point external to the magnetic shell is equal to the product of the intensity of the current into the solid angle which the shell subtends at the given point.

It follows that the work which is done upon a unit north pole by the magnetic forces due to the current is $i(\Omega_1 - \Omega_2)$, if the pole passes from a place where Ω has the value Ω_1 to a place where it has the value Ω_2. In particular, if the pole completely describes a closed path which does not pass through the interior of the circuit in which the current flows, the work is zero. If the pole passes by an external path from a point P, which is infinitely close to one side of the shell, to a point P′, which is infinitely near to P on the opposite side of the shell, Ω changes by an amount which is infinitely nearly equal to 4π, and the work which is done by the magnetic forces is practically $4\pi i$. But *any* shell of the proper moment, which is bounded by the circuit, produces the same magnetic effect at distant points as the current produces; and so we may now replace the shell, whose action we are considering, by another shell of the same moment, which is everywhere finitely distant from P and P′, and which is bounded by the same circuit. The forces due to this shell do infinitely little work upon the unit north pole when it passes from P′ to P, and hence the work which is done upon the unit pole in a single complete passage round a closed path which passes through the circuit is $4\pi i$. In n such passages the work is $4\pi in$. The reason why we must not take account of the work which would be done by the forces inside the shell is simply that the shell is only used for the purpose of representing the action outside the circuit. In the interior of the shell the force due to the shell has necessarily a direction *opposite* to that of the force due to the circular current.

347. *Circular Circuits. Solenoids. Ampère's Hypothesis regarding Magnetism.*—A circular circuit, of radius r, which is traversed by a current i, may be replaced by its equivalent magnetic shell of intensity $\pm i/t$. So also, a second circular circuit, which is traversed by a current i', may be replaced by a shell of intensity $\pm i'/t'$. These shells will, if free, turn so as to present oppositely magnetised faces to each other, and will then exhibit mutual attraction. Hence the electric circuits will tend to turn so

that the currents in each are parallel, and will then exhibit mutual attraction. But if while the circuits are in this position, one of the currents be reversed, mutual repulsion will be exhibited.

[It is an easy matter to calculate the force which such a circuit carrying a current i, exerts at its centre. Let the centre be at o (Fig. 175), and let a, c, be the points in which the circuit cuts the plane of the paper—the plane of the circuit being supposed to be perpendicular to that plane. Imagine the circuit to be replaced by an equivalent hemispherical magnetic shell abc. The work which is done in displacing a unit pole from o through a small distance $od = \tau$ is $i(\omega - \omega')$, where ω and ω' are respectively the angles subtended at o and d by the shell abc. If r be the radius, the value of ω is 2π, and the value of ω' is practically $(2\pi r^2 - 2\pi r\tau)/r^2$; so that the work is $2\pi i\tau/r$. Hence the force, which is practically uniform when τ is sufficiently small, is $2\pi i/r$; so that, when r is unity, unit length of the current exerts a force i at the centre.]

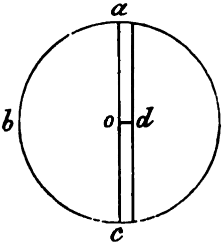


FIG. 175.

These phenomena can be readily exhibited by means of two small floating cells, each of which consists of a test tube containing dilute sulphuric acid into which dip zinc and copper plates connected externally by a circular copper wire. The test tubes are inserted in pieces of cork, and are floated on the surface of water.

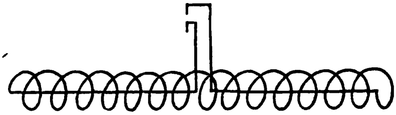


FIG. 176.

A wire which is bent into a cylindrical helix in the manner indicated in Fig. 176, is called a *Solenoid*. If it be freely suspended on pivots, and be traversed by a current, it will act like a magnet under the action of the earth's force or of other magnetic fields. If the number of turns, n, per unit of length is large, we may replace each nearly closed circuit by a shell, the intensity of magnetisation of which is $\pm ni$. For the thickness of each shell equivalent to a turn of the wire is $1/n$, and therefore the surface density of the distribution of magnetism on its two faces is $+ni$ and $-ni$ respectively. Throughout the length of the solenoid, the actions of the shells are mutually annulled, except near the ends, where quantities of magnetism $\pm nia$ are found, a being the area of the shells. At points which are far distant in comparison with the radius of the solenoid, the action is therefore equivalent to that of a magnet of moment $nial$, where l is the length of the solenoid.

In the interior of a very long solenoid, which contains n turns per unit of its length, and through which a current i circulates, the total force is equal to $4\pi nia$, where a is the area of a transverse section. For the force at any point within each shell, and therefore throughout the interior of the solenoid, is $4\pi ni$ (*cf.* § 289). Thus, a solenoid may be used for the production of a strong magnetic field.

The product of current strength, expressed in amperes, into number of turns is called the *ampere-turns* in the solenoid. The force, in absolute units, is, therefore, equal to $4\pi/10$ times the number of ampere-turns per unit of length, since the value of the current strength when expressed in amperes is ten times its value in absolute units (§ 386).

348. *Influence of the Medium. Magnetic and Diamagnetic Media.*—In the previous sections it has been tacitly assumed that the medium in which the various observations under consideration were made was air. If other media be employed, the strength of the field which is produced by a current of given intensity, flowing in a given circuit, may be very different from that which is obtained in air under the same circumstances. It is customary to compare all media with vacuum as the standard. Media which give stronger fields than the standard medium gives are called *magnetic* or sometimes *para-magnetic*, media ; those which give weaker fields are called *diamagnetic* media. When a portion of a medium in which a uniform field of force exists is replaced by a more magnetic medium, the field ceases to be uniform. The lines of force crowd into the more magnetic medium, giving rise to south magnetisation where they enter it and to north magnetisation where they leave it. If a less magnetic medium, or a diamagnetic medium, be introduced, the lines of force are dispersed more,

These phenomena are analogous to those which are exhibited by media of various specific inductive capacities when they are subjected to electrostatic stress.

849. *Material Magnets.*—Some substances retain, to a greater or less extent, their state of magnetisation after the magnetising force is removed, and so produce a more or less permanent magnetic field in their neighbourhood, *i.e.*, continue to act more or less as magnets. The property in virtue of which this occurs is termed *Retentiveness.*

The magnetism which remains, because of retentiveness, is called *Residual Magnetism.* From its great retentiveness, hard steel is employed in the construction of so-called permanent magnets. The residual magnetism of a long bar of steel is more permanent than is that of a short bar, for the self demagnetising force which such a bar necessarily exhibits (§ 852) has less influence in the former case than it has in the latter.

Retentiveness has very different values in different materials. It is relatively small in good specimens of soft iron.

In order to get rid of residual magnetism in any substance, we must either heat the substance to redness, or employ a reverse magnetising force. It is usual, therefore, to speak of a *Coercive Force*, existent in the material, in virtue of which residual magnetism is retained.

The property of permanent magnetisation is possessed notably by one of the oxides of iron (lodestone, found abundantly in Magnesia —hence the name *magnet*) in its natural state, but it may be induced to a very much greater extent in pieces of steel or metallic iron. The metals cobalt and nickel are also capable of becoming strongly magnetic. All other substances have relatively extremely feeble magnetic properties.

The most powerful magnets are *electro-magnets.* In such a magnet an electric current flows round a solenoid or coil of insulated wire which is wound round a core of soft iron.

No ordinary material magnet completely realises the conditions which are supposed to exist in an ideal magnet. It is practically impossible to uniformly magnetise, *e.g.*, a 'bar' magnet—a magnet made of a rectangular or cylindrical bar of steel. Yet, at any one external point, such a magnet really acts as a certain definite ideal magnet would act, though the equivalent ideal magnet is in general quite different when the action at another external point is considered. It is more convenient therefore to speak of the north and south 'ends' of such a magnet rather than of its north and south 'poles.'

When the external points are not too close to the magnet in comparison with its dimensions, the magnet practically acts as an ideal magnet with definite fixed poles. Then too—and only then—the magnet possesses a definite magnetic moment.

Such magnets may be conveniently used to test the laws of magnetic attraction and repulsion, and to exhibit the phenomena of induced magnetisation in magnetisable bodies—phenomena which take place in close analogy to those of electrostatic induction.

350. *Magnetism a Molecular Phenomenon in Material Media.*—The great distinction between electrical and magnetic phenomena lies in the absence of anything of the nature of *conduction* of magnetism. While it might seem that the disappearance of induced magnetisation when a piece of soft iron is withdrawn from the neighbourhood of a magnet is due to the flowing together of the two opposite kinds of magnetism, the persistence of magnetisation to an appreciable extent when the soft iron is replaced by hard steel at once disposes of this view.

And, further, if we bring a magnetic substance into contact with one end of a magnet and then withdraw it from contact, no interchange of magnetism takes place, though an interchange of electricity would occur if the substances were electrified conductors. Also, while a conductor under the influence of an electrified body may be divided into two oppositely charged portions, it is impossible to divide a magnet into two oppositely magnetised portions—that is to say, it is impossible to isolate one kind of magnetism.

Every portion, however small it may be, into which a magnet may be broken exhibits properties precisely similar to those which were manifested by the complete magnet. We conclude, therefore, that this would still hold if the magnet were reduced to its constituent molecules; that each molecule of a magnetised body is itself a little magnet.

It is easy to explain, upon this assumption, how it is that magnetisation is only evident near the ends of a magnet. For, if the little circles in Fig. 177 represent the magnetised molecules, we see

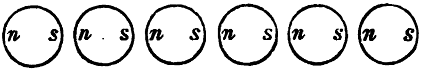


FIG. 177.

that the effect of any north end of a molecule at external points is counterbalanced by the effect of the south end of an immediately adjacent molecule. It is only at the extremities of such a chain of molecules that the magnetisation can become manifest through the production of external effects, and the magnetism is of opposite kinds at the two extremities of the chain.

351. *Experimental Mapping of Lines of Magnetic Force.*—If we draw from any magnetic pole a number of lines of force numerically equal to 4π times the strength of the pole, the number of these which cross unit area of any plane surface passing through any point can be made to represent the strength of the field—*i.e.*, the magnitude of the force—at that point in the direction of the normal to the given plane. In fact, as we have already seen, all the results previously given regarding electric lines of force can be at once applied to magnetic lines of force, and we, therefore, do not require to repeat them here. It is merely necessary to replace the term 'electrified body' by the term 'magnetised body,' the term 'positive electricity' by the term 'north magnetism,' 'negative charge' by 'quantity of south magnetism,' and so on.

A line of force can be readily traced out by means of a very small magnet, freely suspended, which is always moved in the direction in which it points. In every position its length is tangential to the line of force which passes through its centre. The lines of force due to any group of magnets can also be readily shown by means of iron filings dusted over a sheet of paper, which is placed over the magnets. The filings become magnetised and turn so as to place their lengths in the direction of the force. A slight lapping of the paper will cause the filings to group themselves in

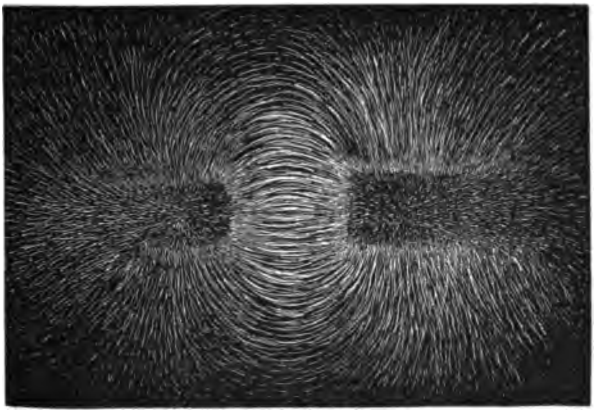


FIG. 178.

definite lines, each of which coincides with a line of force. The vibration of the paper throws the filings up into the air for a moment, so that they are free to accommodate themselves to the influence of neighbouring filings. (See Figs. 178, 179.)

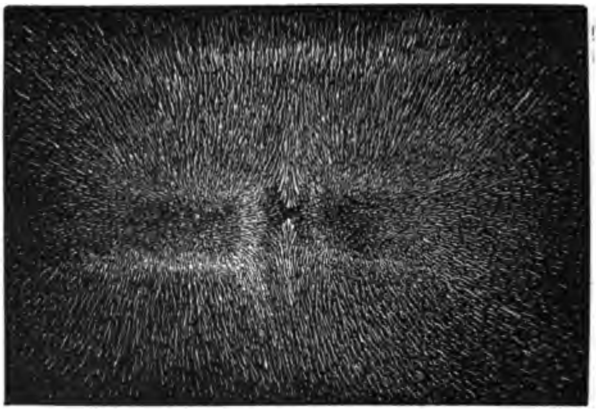


FIG. 179.

852. *Magnetic Induction.*—If a rectangular bar-magnet be uniformly magnetised in the direction of its length (say), it is obvious that its total magnetic moment is equal to the sum of the moments of any number of parts (uniformly magnetised in the same manner), into which we may suppose it to be divided. For, if $L = l_1 + l_2 + \ldots\ldots + l_n$ be the total length of the magnet, and if Q be its pole-strength, we have

$$LQ = (l_1 + l_2 + \ldots\ldots + l_n)Q = l_1Q + l_2Q + \ldots\ldots + l_nQ,$$

which proves the proposition so far as transverse division is concerned. And, since the magnetisation is uniform, if we divide it longitudinally, each part becomes a magnet, whose pole-strength is proportional to the area of its end. Hence, if $A = a_1 + a_2 + \ldots\ldots a_n$ be the total area of the end, and if $\mathfrak{I}$ be the strength per unit of area, so that $\mathfrak{I}A = Q = \mathfrak{I}(a_1 + a_2 + \ldots\ldots + a_n) = q_1 + q_2 + \ldots\ldots + q_n$, where q_1, etc., are the strengths of the several parts, we have

$$LQ = Lq_1 + Lq_2 + \ldots\ldots + Lq_n,$$

which proves the statement for longitudinal division.

But it is obvious that the statement is true whatever be the forms of the parts into which the magnet is divided, for each part may be supposed to be built up of an infinitely great number of infinitely small rectangular portions. And it follows from this consideration also that the proposition is true whatever be the form of the original magnet.

The quantity $\mathfrak{I}$, the pole-strength per unit of area, is identical with the *Intensity of Magnetisation* as previously defined. It is evident that we may regard it as being the magnetic moment per unit of volume.

Let us imagine a cylindrical crevasse to be cut out in the interior of a uniformly magnetised body. Let it be bounded by plane surfaces perpendicular to the direction of magnetisation; and, while all its dimensions are infinitely small, let the perpendicular distance between these planes be infinitely smaller than the transverse dimensions. The surface density of magnetism on the plane faces of the cavity is $\mathfrak{I}$, north magnetism being distributed on the plane face next the south pole of the magnet, while south magnetism is distributed on the other; and hence the force in the space between the planes is $4\pi\mathfrak{I}$ (§ 289). We may therefore suppose that $4\pi\mathfrak{I}$ lines of force are drawn per unit of area across this cavity in the direction of magnetisation (that is, from the south pole to the north pole within the substance of the magnet). It is customary to call these lines the *Lines of Magnetisation.*

The lines of magnetisation do not constitute all the lines of force in the interior of the magnet. There may be lines of force due to external magnetisation. This distribution of force must be investigated in precisely the same manner as that in which we investigate the distribution of force outside a magnet.

Let us suppose that the cylindrical cavity is infinitely long in comparison with its cross dimensions. The magnetisation at the ends of this cavity exerts no effect upon a point at its centre, and hence any force found at the centre must be due to external magnetisation. This quantity is denoted by the symbol $\mathfrak{H}$, and is called the magnetic force at the point. The total force, $\mathfrak{B}$, is called the *Magnetic Induction* at the given point in the magnet, and is equal to

$$4\pi\mathfrak{I}+\mathfrak{H}.$$

It is usual to call the lines of total force inside a magnet, *Lines of Induction.* They consist, therefore, partly of lines of magnetisation, and partly of the lines of force within the uncut magnet. They are continuous with the lines of force external to the magnet.

[It must be remembered that the three quantities $\mathfrak{B}$, $\mathfrak{I}$, and $\mathfrak{H}$, are vector quantities, and are therefore subject to the laws of vector addition (§ 24). In most practical cases, however, $\mathfrak{I}$ and $\mathfrak{H}$ are similarly directed.]

The force due to the magnetism at the surface of the magnetised body is included in the quantity $\mathfrak{H}$, and is obviously directed oppositely to $\mathfrak{I}$ since it acts in the direction of a line drawn from the north pole to the south pole through the material of the body. It therefore acts so as to demagnetise the body, and has its greatest value $2\pi\mathfrak{I}$ at points close to the ends of the magnet (compare § 289). [To obviate demagnetisation, bar magnets, when not in use, are placed parallel to each other with their like poles oppositely directed, and 'keepers' made of a magnetic metal (soft iron preferably) are placed in contact with their ends (Fig. 180). A closed magnetic circuit is thus formed, for the effect of the magnetism at the poles is annulled by that of the magnetism which is induced in the keepers.]

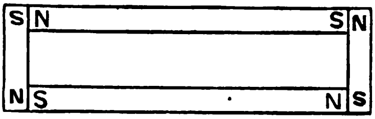


FIG. 180.

It appears, therefore, that the form of a magnet must have an effect upon the distribution of magnetisation throughout its interior. Thus, while in a long rod placed parallel in the direction of the force in a uniform field, the magnetisation is sensibly uniform except near the ends, in a short rod the magnetisation is very far from being uniform.

353. *Permeability and Susceptibility.*—That property of a substance in virtue of which the lines of induction are more or less closely arranged than are the lines of force in the originally undisturbed field is called the *Permeability* of the substance. We have

$$\frac{\mathfrak{B}}{\mathfrak{H}} = 4\pi\frac{\mathfrak{I}}{\mathfrak{H}} + 1,$$

which may be written in the form

$$\mu = 4\pi k + 1.$$

In this equation, the quantity μ represents the permeability, and k represents the *Susceptibility*. The permeability is therefore the ratio of the induction to the force within the substance of the magnetic body, while the susceptibility is the ratio of the magnetisation to the magnetising force, and is a measure of the readiness of the body to acquire magnetisation.

In a paramagnetic substance, as we have already seen, the lines of induction are more closely arranged than are the lines of force. That is to say, the permeability of such a substance is greater than unity; and therefore the susceptibility is positive. On the other hand, in a diamagnetic substance, the lines of induction are less closely arranged than are the lines of force; and so the permeability is less than unity, and the susceptibility is negative. In a paramagnetic body, north magnetism is manifested at that extremity which faces in the direction in which the external lines of force are drawn; in a diamagnetic body, north magnetism appears at the opposite end. Consequently, while a paramagnetic substance is attracted towards the pole of a magnet, a diamagnetic substance is repelled from it. In Fig. 181, the body marked *p* is paramagnetic, the body marked *d* is diamagnetic.

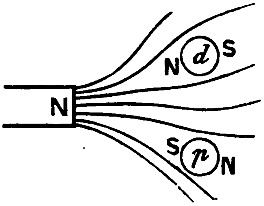


Fig. 181.

More generally—a paramagnetic substance moves from weak parts to strong parts of a field of force, while a diamagnetic substance moves from strong parts to weak parts.

It must not be supposed that the appearance of north magnetism at the face of the diamagnetic body which is near the north end of a magnet implies that the body exhibits reverse magnetisation. It is more strictly correct to say that the north magnetism is developed on the face of the intervening medium.

354. *Law of Magnetic Circuits.* – From the various results which have been considered, it appears that the total lines of magnetic force present close analogies to the lines of electrostatic force in dielectric media. The special use which has been made of the term 'lines of magnetic induction' is not altogether free from objection, since the analogous term regarding electrical action has a different meaning. And, strictly speaking, if the use of the term is retained, the external lines of force, with which the total internal lines are continuous, should be called the lines of magnetic induction in the external medium. The term *Magnetic Flux* is therefore often used in the definite sense of total magnetic induction, although

it is not intrinsically a more desirable term. The total force which is employed in air—more correctly, in vacuo—to produce magnetic flux (say by means of a solenoid) is called the *Magnetomotive Force.* From the investigation in §§ 346, 347, it is evident that this quantity is equal to the work which is done in driving a unit north pole once round the complete circuit of a line of force.

Other things being equal, the magnetic flux is proportional to the magnetomotive force. In different media, it is also proportional to the permeabilities. With regard to a given magnetic circuit, we may employ the term *Permeance* so that permeance and permeability correspond to conductance and conductivity in electrical conducting circuits, or, more correctly, to inductive capacity and specific inductive capacity in dielectrics. The law of magnetic flux may then be expressed so as to present a complete analogy to the law of electrical conduction or to that of electric displacement in dielectrics. Intensity of current is proportional to the product of electromotive force into conductance; electric displacement is proportional to the product of electric force into inductive capacity; and, similarly, magnetic flux is proportional to the product of magnetomotive force and permeance. It is more customary to employ the reciprocal of the permeance, called the *Reluctance*, and to write

$$\mathrm{M}=\mathrm{NR},$$

where M is the magnetomotive force, N is the magnetic flux or number of magnetic lines, and R is the reluctance.

The total reluctance of a magnetic circuit is the sum of the reluctances of its several parts. The reluctance of any portion is equal to its specific reluctance or *reluctivity*—the reciprocal of its permeability—multiplied by its length and divided by its sectional area. Hence if l, a, and μ, with distinctive suffixes, represent respectively length, sectional area, and the permeability corresponding to the given value of the magnetising force (§ 355),

$$10\mathrm{N}\left(\frac{l_1}{a_1\mu_1}+\frac{l_1}{a_1\mu_1}+\ldots\right)=4\pi n,$$

when n represents the total number of ampere turns in the magnetising circuit.

355. *Relation connecting Magnetisation and Magnetising Force.*—If the magnitudes of any two of the quantities $\mathfrak{B}$, $\mathfrak{I}$, and $\mathfrak{H}$, are determined in any particular case, the value of the remaining quantity can be calculated from the relation $\mathfrak{B}=4\pi\mathfrak{I}+\mathfrak{H}$. Methods for the determination of each of the three quantities will be described subsequently.

Fig. 182 represents the usual course of the variation of intensity with magnetising force. The force is measured along the axis O$\mathfrak{H}$, while the intensity of magnetisation is measured along the axis O$\mathfrak{I}$.

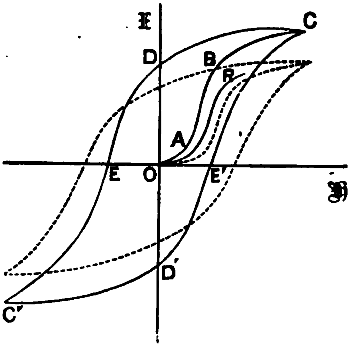


FIG. 182.

At first, while the force is small, the magnetisation increases very slowly, and at a sensibly uniform rate. Then, as the force is increased, the law becomes $\mathfrak{I}=a\mathfrak{H}+b\mathfrak{H}^2$, a and b being constants. After this, a very slight increase of the magnetising force produces a great change in the magnetisation. With still larger forces, the rate of variation becomes rapidly smaller, and ultimately the magnetisation becomes sensibly constant. These various stages in the process of magnetisation are represented by the parts OA, AB, and BC, of the curve. If the force be now gradually removed, the magnetisation will diminish at a relatively slow rate, until, when the force is entirely removed, a considerable amount of residual magnetisation remains. This is represented by OD.

If a reverse force be now applied, the magnetisation will fall off rapidly in magnitude, and will disappear entirely when the reverse force has a definite value OE. This may be supposed to represent, as Hopkinson suggests, the coercive force.

If the reverse force be now increased until it reaches a value equal to the maximum value of the direct force, if it be then diminished to zero, and if, finally, positive force be reapplied until the original maximum value is attained, the magnetisation will pass through successive values represented by the part EC′D′E′C..

The curve OR represents the residual magnetisation which is left after various magnetising forces have been applied and removed.

The dotted curve represents the change which takes place in the magnetisation when the same substance (say a soft iron wire) is hardened by being stretched beyond its limits of elasticity. The maximum magnetisation is lessened. The residual magnetisation is also lessened, but the coercive force is increased.

Since the magnetisation practically reaches a maximum value when the force is sufficiently great, the substance is then said to be *saturated.* However much the force may be further increased, the intensity remains appreciably constant.

The susceptibility increases from a small value to a maximum which is indicated by a tangent drawn from O to the curve OBC. Thereafter it diminishes to zero as the force increases without limit. The relation $\mu=4\pi k+1$ shows that the permeability also increases from a small value to a maximum (attained at a somewhat greater value of $\mathfrak{H}$ than that at which the maximum susceptibility is reached), after which it gradually diminishes to unity as the force is indefinitely increased.

The following table gives, in the first and second columns respectively, values of $\mathfrak{H}$ and μ given by Hopkinson for annealed wrought iron.

1·66	3000
4·00	2250
5·00	2000
6·50	1692
8·50	1412
12·00	1083
17·00	823
28·50	526
50·00	320
105·00	161
200·00	90
350·00	54
666·00	30

From these numbers the values of the susceptibility, the intensity of magnetisation, and the total induction or flux, may be calculated.

When soft iron is magnetised (more especially when the force is feeble and the specimen of iron is large) it is found that the magnetisation takes some time to attain the full value corresponding to the force which is acting. This effect is, by analogy, said to be due to *Magnetic Viscosity.*

356. *Hysterēsis.*—We see from Fig. 182, that the changes of magnetisation tend to lag behind the changes of force which give rise to them. Thus, when the stage C has been reached, a much greater change in the value of the force is requisite in order to effect a given diminution in the magnetisation than was requisite for the production of an equal increase of magnetisation just before the stage C was reached. A similar effect is observable at the point C′. Ewing has called this tendency *Magnetic Hysterēsis.*

As the result of hysterēsis, different values of the magnetisation may correspond to one given value of the magnetic force, and we must therefore limit our definitions of permeability and susceptibility to the case of a substance which is originally unmagnetised and which is subjected to a force which increases in magnitude continuously from zero upwards.

If e represent the increase of the magnetic energy of the magnetised body, which accompanies a small increase of intensity of magnetisation, i, the quantity e/i is (§ 48) the force which produces the change i: that is, it is the force H. Hence the increment of energy per unit of volume which accompanies the increment of i is $\mathrm{H}i$; and therefore the area CDC′D′C (Fig. 182) represents an amount of energy which has been transformed, per unit of volume, in the given cyclical process. This energy takes the form of heat, and is dissipated. Consequently, rapid reversals of magnetism, will cause the temperature of the magnetised substance to increase markedly; and no amount of lamination, such as is used in transformers or in the armature cores of dynamos for the prevention of heating by induced currents, will prevent this effect.

No dissipation of energy occurs if the cyclical changes in the magnetising force are small and take place either very rapidly or very slowly. For, in the former case, no time is allowed for a diminution to occur in the amount of lag of the magnetic effect behind the change of force which produces it, whether in the direct or in the reverse part of the cycle; so that the direct changes of magnetisation are exactly reversed when the force is reversed; and, in the latter case, complete time is allowed to prevent noticeable lag from making its appearance, that is to say, the changes in the force take place so slowly that the proper changes of magnetisation can ensue at all stages of the process; and so, again, the reverse changes of magnetisation follow, in the opposite direction, the same course as the direct changes. In any other case dissipation of energy will be manifested.

It appears, therefore, that the so-called magnetic viscosity tends to produce hysterēsis. But, though this is so, the converse statement that the existence of hysterēsis implies the existence of viscosity is neither necessarily nor actually true.

357. *Effects of Vibration and of Temperature.*—Vibration has a very great influence upon the susceptibility of a magnetised body. This effect is very marked when the magnetising force is small, but is not very noticeable, if at all, when powerful forces are used. It increases the susceptibility of the substance, but diminishes residual magnetism, coercive force, and hysterēsis. These results are ex-

hibited in Fig. 183, in which the full curve represents a cycle performed under the condition of no vibration; while the dotted curve represents the result of an experiment made upon the same

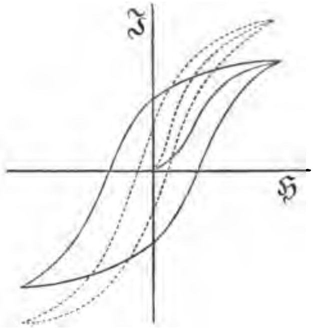


Fig. 183.

substance under similar conditions of force—the substances, however, being tapped after each change in the magnitude of the force.

The temperature of a magnetic substance, too, has a very marked effect upon its susceptibility. In iron, cobalt, and nickel, increase of temperature (from ordinary values) first increases the susceptibility, and afterwards diminishes it, as the magnetising force is continuously increased; and the magnetic properties entirely, and suddenly, vanish when the temperature attains a certain value which is different for each substance, and varies to some extent also from one to another specimen of any substance.

The temperature at which the susceptibility vanishes is called the *Critical Temperature.* It is a temperature at which some fundamental change takes place in the physical constitution of the metal. The electric resistance of this metal changes suddenly at this point, as also does its thermo-electric power (§ 301). When the reverse change—from the non-magnetic condition to the magnetic condition—takes place, as hard steel is cooled down from a temperature higher than the critical temperature, a sudden liberation of heat takes place, and the metal glows brightly, although it had previously cooled to dull redness.

In iron, the suddenness with which the magnetisation is lost as the critical temperature is approached, depends very largely upon the value of the magnetising force. When the force is very small, the susceptibility first increases with extreme rapidity to a maximum, and then diminishes with even greater rapidity, as the critical temperature is approached. With higher forces, the variation becomes much less marked.

There is little or no evidence of hysterĕsis with regard to the magnetic effects which follow changes of temperature unless the critical temperature be included within the cyclical range. But it at once becomes evident when the range includes the critical temperature; for the temperature at which the magnetic effects reappear as the temperature is reduced, is lower than the critical temperature at which they disappear when the temperature is raised.

This lag of magnetic effect, behind the change of temperature which gives rise to it, is abnormally evident in certain alloys of nickel and iron. An alloy containing 25 per cent. of nickel was found by Hopkinson to lose its magnetic properties at a temperature of 580° C., and to remain non-magnetic until its temperature fell somewhat below the Centigrade zero. This fact suggests the idea that non-magnetic manganese steel may become magnetic if its temperature be sufficiently reduced, and that possibly all the non-magnetic metals may act similarly.

358. *Effects of Stress.*—Alteration of the state of stress to which the magnetic metals are subjected produces considerable alteration of the magnetic qualities of the metals.

Matteuci observed that extension of an iron rod produced an increase of magnetisation, and Villari found that when the field is sufficiently intense, extension causes decrease of magnetisation. This effect is called the 'Villari reversal.'

The various effects of longitudinal and of torsional stress have been very fully investigated by Wiedemann, Lord Kelvin, and others.

Compression of an iron rod produces effects opposite to those which are produced by extension. Compression and extension, respectively, of nickel and cobalt rods also produce respectively opposite effects, but there is no Villari reversal in this case; for all values of the magnetising force, extension produces diminution, and compression produces increase, of the magnetisation. The diminution of the residual magnetisation of nickel, under extending stress, is even more evident than is the diminution of induced magnetisation. Hysterĕsis, under cyclical variation of load, is much more marked in the case of iron than it is in the case of nickel.

From the above result regarding the effect of extension on the magnetisation of an iron rod in a weak field, we can, by a double application of the principle of stable equilibrium (§ 15), deduce the result that, in weak fields, increase of magnetisation causes increase of length; or, conversely, we can deduce the former result from the latter. Magnetic energy enters the rod from the external medium, and, in part, is transformed into potential energy of molecular configuration within the rod; and this potential energy may be, in turn, transformed into external work as the length of the rod alters. Let us suppose, first, that the length of the rod is not allowed to alter. Increase of magnetisation will then give rise to pressure on the restraining surfaces which prevent alteration of length. Conversely, by the principle of stable equilibrium, decrease of pressure will cause an increase of magnetisation under the given external magnetic force. But, again, increase of pressure upon the restraining surfaces will result in increase of length of the rod if the restraint be removed. Conversely, increase of length will cause a diminution of pressure. We may represent these results by the symbolical expression

$$\begin{tikzcd} +M \arrow[r] & +P \arrow[r] & +L \arrow[d, equal] \\ +M & -P \arrow[l] & +L \arrow[l] \end{tikzcd}$$

where M, P, and L represent respectively magnetisation in weak fields, pressure, and length. Disregarding the intermediate step, the symbols state that increase of the magnetisation of an iron rod in weak fields causes increase of length, and that increase of length of the rod induces an increase of magnetisation. In strong fields the expression would become

$$\begin{tikzcd} +M \arrow[r] & -P \arrow[r] & -L \arrow[d, equal] \\ +M & +P \arrow[l] & -L \arrow[l] \end{tikzcd}$$

where −P may be translated 'increase of tension.'

[It is important to observe that, when we regard only changes of magnetisation and of pressure (or tension), we are dealing with energy flowing from an external system into the iron; and that, when we regard only changes of pressure (or tension) and of length, we are dealing with energy flowing from the iron rod to another external system; while, when we regard changes of magnetisation and of length alone, we are dealing with energy flowing through the rod from one external system to another, and being transformed through its agency from one form to another.]

Joule proved that no observable change of volume takes place when a rod of iron is magnetised, and therefore that longitudinal magnetisation in weak fields must cause a diminution of the sectional area of the rod. Hence he concluded that if a rod of iron be magnetised circularly, that is, if the lines of magnetisation be circles surrounding the axis of the rod, longitudinal contraction will ensue. He verified this conclusion by experiment.

Torsional strain, also, is accompanied by variations in the magnetic qualities of iron, nickel, and cobalt rods. These effects can, as Lord Kelvin has shown, be deduced from the known effects of longitudinal stress upon the magnetic qualities. Thus it is known that the susceptibility of iron in weak fields is increased along lines of traction, and is decreased along lines of compression. But, when a circular rod of iron is twisted in the manner which is indicated by the arrows in Fig. 184, all lines such as *a a′* suffer traction, while all lines such as *b b′* suffer compression. Hence the susceptibility is increased along *a a′* and is diminished along *b b′*. The effect of this is practically to produce two components of magnetisation—one longitudinal, the other circular—when the twist is sufficiently great. Hence torsion in weak fields diminishes the longitudinal susceptibility of iron.

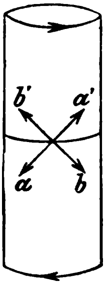


FIG. 184.

Conversely, a circularly magnetised iron rod, when twisted, becomes longitudinally magnetised. It is easy to deduce the corresponding effects in nickel and cobalt. [No reversal of the direction of longitudinal magnetisation takes place in iron, however strong the circular magnetisation may be. Ewing explains this by the fact that the intensity of magnetisation in the direction of the line of traction, or of compression, never reaches the point at which the Villari reversal occurs.]

Since torsional stress produces circular magnetisation in a longitudinally magnetised rod, and since it also gives rise to longitudinal magnetisation in a circularly magnetised rod, we might expect that the superposition of longitudinal and circular magnetisations would cause torsional strain. This effect was discovered experimentally by Wiedemann in the case of iron. The twist in weak fields takes place in such a direction as to be completely explainable by the increase of length which occurs in the direction of resultant magnetisation. Knott has shown that the twist occurs in the reverse way in nickel—a result which he expected to find, since nickel contracts in the direction of magnetisation.

We may observe here that the twisting of a magnetised rod, or the magnetisation of a twisted rod, gives rise to a transient electric current in the magnetised material. This effect will be considered in the next chapter.

359. *Magnetometric Measurements.*—The *magnetometer* consists essentially of a small magnet, which is suspended by a long fine fibre, whose co-efficient of torsion is negligable in most cases, and which is free to turn about that fibre as an axis. A small mirror is usually attached to the magnet, so that, by means of a reflected beam of light, very small angular motions of the magnet may be made evident. This apparatus is placed in a uniform field of force of known intensity, say the earth's field (§ 360). The magnet then places its length in the direction of the force in the given field.

Let the magnet be placed at P (Fig. 185), and let the direction of the controlling force be PQ. Let AB be a bar magnet, the intensity of magnetisation of which we have to determine, and let the points A and B represent the position of its poles. Place it symmetrically with regard to PQ in the position which is indicated in the figure. Let I be its (unknown) intensity of magnetisation, while a is its sectional area. Then Ia is the strength of its pole. The effect of the north pole, A, at P is in the direction AP, and is equal to

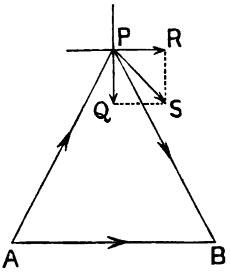


FIG. 185.

Ia/AP^2. Similarly, the directive force of the south pole, B, at P is in the direction PB, and is equal to $-$Ia/PB^2. If we represent these forces by AP and PB respectively, it is obvious that the resultant effect of the two is represented on the same scale by AB. The magnitude of the resultant is therefore $IaAB/AP^3$, and acts so as to place the little magnet at P parallel to AB, with its poles facing oppositely to those of AB.

Now let PR represent this force on the same scale that PQ represents the force of the external field. PS is the resultant of these forces, and the little magnet sets itself in the direction PS, making an angle θ with PQ, such that $\tan\theta = PR/PQ$. If F be the intensity of the external force, this gives

$$\frac{IaAB}{AP^3} = F\tan\theta, \ldots\ldots\ldots (1),$$

from which we can calculate I.

In order to find the value of the force F, if it be unknown, we may set the magnet AB oscillating under the action of F alone. The time, T, of a small oscillation is given by the equation

$$\frac{4\pi^2}{T^2} = \frac{FIaB}{K},$$

where K is the moment of inertia of the magnet about its axis of suspension, and AB is its length. For if the direction of the force F be denoted by the arrow (Fig. 186), and if *ns* represent the magnet inclined at an angle θ to the direction of the force, F sin θ is the force which is acting perpendicularly to the length of the magnet, and which tends to turn it around its axis of suspension. This force acts at each end of the magnet so as to produce rotation in the positive direction, and the turning moment is therefore

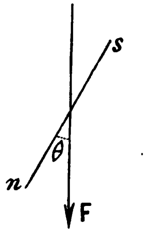


FIG. 186.

F sin θ $IaAB$. When the angle is small, this becomes $F\theta IaAB$. If the angular acceleration be α, the momentum which is produced per unit of time is $m\alpha R$, where m is the mass of the magnet and R is its radius of gyration (§ 53). Hence the moment of momentum which is produced per unit of time is $ma\text{R}^2 = \text{K}a$, where K is the moment of inertia (§ 53). We therefore have

$$\text{K}a = -\text{F}\theta\text{I}a\text{AB},$$

the minus sign being used, since the angular acceleration is negative. Now every quantity in this equation is constant, with the exception of θ; and so the equation expresses the fact that the angular acceleration is negative and is proportional to the displacement. The small oscillations of the magnet, therefore, obey the simple harmonic law, and the angular position of the magnet is given (§ 33) by the equation

$$\theta = \text{P} \cos\left(\sqrt{\frac{\text{FI}a\text{AB}}{\text{K}}} \,.\, t + \text{Q}\right),$$

where P and Q are constants and t is the time; whence T being the periodic time, we get (§ 33)

$$\frac{4\pi^2}{\text{T}^2} = \frac{\text{FI}a\text{AB}}{\text{K}} \,\ldots\ldots\ldots\, (2).$$

By elimination between the equations (1) and (2), we can find the values of F and of IaAB (which is the magnetic moment of the magnet). Also, a and AB being known, we can obtain the value of I; and hence, if we know the intensity of the magnetising force, we can calculate the susceptibility and the permeability of the substance.

Equation (2) shows that the intensities of two fields are inversely proportional to the squares of the periods of the small oscillations of a magnet of known magnetic moment, which is suspended first in one field and then in the other.

The ballistic method of making magnetic measurements will be discussed in § 367.

360. *Terrestrial Magnetism.*—The earth exerts a magnetic action in virtue of which compass needles point in a northerly direction. The angular distance between the line along which a compass needle points and the geographical meridian is called the magnetic *declination* or *variation.* The magnetic needle, if it were carefully supported on a horizontal axis which passes through its centre of inertia, and which is perpendicular to the magnetic meridian, would, in Britain, place its magnetic axis in a direction which is inclined to the horizon—the north end pointing downwards at a considerable angle. This angle is called the *Magnetic Dip.*

The declination and the dip vary considerably from one part of the earth's surface to another. In some regions the declination is easterly, in others it is westerly. The line on the earth's surface, at all points of which the declination is zero, is called the *Magnetic Equator.* It does not coincide with the geographical equator, and is not a great circle. That point on the surface at which the north pole of a magnet points vertically downwards is called the *North Magnetic Pole* of the earth, and that point at which the south pole of a magnet points vertically downwards is called the *South Magnetic Pole* of the earth. These poles do not coincide with the geographical poles of the earth, neither do they lie at opposite extremities of a diameter. [Observe that the magnetism which we may suppose to be collected at the *north* pole of the earth must be *south* magnetism, *i.e.*, it must be of the same kind as that which appears at the south-pointing pole of a magnet. Similarly, the magnetism which is manifested at the south magnetic pole of the earth must be north magnetism.]

The earth's magnetic force is in a constant state of variation. It changes with the hour of the day and the time of the year; and it depends also upon the position of the moon. Yet these variations do not appear to be due to any *direct* action of the sun or the moon.

Sudden disturbances sometimes take place in addition to these more regular variations. A period of maximum disturbance occurs every eleven years, and coincides with the period of maximum sun-spot disturbance.

A slow secular change of the position of the magnetic poles is also in progress.

S. P. Thompson suggests that variations of the earth's magnetism are due to convection currents of positive electricity, carried northwards and southwards by the heated air from the equators, inside which the earth rotates. The return current may in large part constitute the 'earth-currents' which flow in the material of the earth and are shunted through telegraph-wires. On the other hand, it may be the case that the electric currents of the upper atmosphere, which presumably cause the auroræ, are due to the variations of the earth's magnetism.

861. *Theories of Magnetism.*—At one time magnetic phenomena were explained by the assumption of the existence of two imponderable fluids, one of which constituted north magnetism, while the other constituted south magnetism. In Poisson's elaboration of this theory a magnetic body was supposed to be made up of spheres of infinite permeability, uniformly distributed in an absolutely non-permeable fluid. This made the problem of magnetic induction identical with that of electric induction in a non-conducting medium throughout which perfectly conducting insulated spheres were uniformly distributed. Among other objections to this theory is the fatal one pointed out by Maxwell, that the permeability of

iron is too great to be accounted for even if the spheres were packed in the closest possible arrangement.

In modern theories the molecules are supposed to be little magnets. In an unmagnetised body the magnetic molecules have their axes distributed, on the whole, uniformly in all directions; and the substance becomes magnetised when the axes of its molecules get, on the whole, a definite set in one direction. Saturation will take place when all the molecules have set their axes in the direction of the magnetising force.

The fact that the slightest force does not produce saturation shows that displacement of the molecules must be resisted by some force. Weber assumed that each molecule is acted upon by a constant force in the original direction of its axis, which tends to prevent its orientation from that direction, and tends to make it resume its original direction when it is displaced from it. It follows from this assumption that the curve of magnetisation (Fig. 182) should at first be a straight line, that it should afterwards become concave to the axis along which the force is measured, and that it should ultimately approach an asymptote parallel to that axis. This does not agree with experiment.

Maxwell improved this hypothesis by the additional assumption that a molecule could return to its original position if it were turned through an angle of less than a certain finite magnitude, and that if it were displaced through an angle greater than this, it would retain, after removal of the force, a displacement equal to the excess of its total displacement over this quantity. This form of the theory leads to a magnetisation curve similar to that given by Weber's unmodified theory, and it indicates that the curve of residual magnetisation starts from a point on the force-axis at a finite distance from the origin, is always concave to that axis, and approaches a parallel asymptote. These results also are incorrect.

Ewing, following a hint by Maxwell, regards each molecule as subject only to the mutual action of the entire system of surrounding molecules. He has constructed a model of such a system by means of a number of pivoted magnets, which are arranged in parallel rows. So long as no external magnetic force acts, the magnets arrange themselves in positions of stable equilibrium under their mutual forces, some of them pointing in one direction, some in another. This illustrates the condition of non-magnetised steel. If only a feeble uniform magnetic force acts, each magnet is slightly turned from its first position, which it reassumes when the force is removed. This illustrates the first stage in the process of magnetisation. A somewhat stronger force causes instability in the originally less stable groups of the magnets, and the magnets which compose these groups swing round into a new stable position. As the external force is increased still further, more and more groups break up, until all have taken the new position of equilibrium under their own mutual forces and the external directive force. This illustrates the second stage of magnetisation, in which the ratio of magnetisation to magnetising force increases with great rapidity. The third stage, in which this ratio is practically constant, is exemplified by the fact that an infinite force is now needed to make the magnets point exactly in the direction of the external lines of force. If the external force be now removed, a considerable proportion of the magnets retain their final positions of equilibrium—in other words, magnetic retentiveness is exhibited.

This model can also show the effects of strain on the magnetic properties. For this purpose the magnets are placed on a sheet of indiarubber. If the indiarubber is stretched the magnets are separated out from one another in one direction, and are brought nearer to each other in a direction at right angles to the former. The magnetic susceptibility is increased or is diminished, according as the stability of the magnets is diminished or is increased by the alteration of relative position. Similarly, the increase of the susceptibility of iron with rise of temperature is explained by the diminution of mutual magnetic influence which results from increased distance. Professor Ewing suggests that the total loss of magnetisation which occurs at a high temperature is due to a continuous whirling motion of the magnetic molecules. He suggests, also, that the dissipation of energy, which occurs when hysterĕsis is exhibited, is due to the induced electric currents (§ 363), which are caused by angular motions of the magnetic molecules.

The magnetic properties of circular circuits in which electric currents flow led Ampère to suggest that the molecules of magnetic substances may exhibit their magnetic properties in virtue of electric currents which circulate within them in closed circuits.

CHAPTER XXIX.

ELECTRO-MAGNETIC INDUCTION. ELECTRO-DYNAMICS.

362. *Action of an external Magnetic Field on an Electric Circuit.*—Let the intensity of magnetisation of a shell, which produces the same magnetic action as a given circuit, be I; and let the shell be placed, in its position of equilibrium, in a field of force the intensity of which is F. We may choose the shell so that each element of its surface is at right angles to the direction of the force in its immediate neighbourhood. Consider a small portion σ of the north face of the surface. The total force acting on this part is $+\mathrm{FI}\sigma$—the plus sign being used since we consider the force to be positive when it acts in the direction of the outward normal to the north face. The increase of the potential energy of any portion, due to the given field, is equal to the work which has been done *against* the forces in displacing the north face forwards, from coincidence with the south face, through the thickness of the shell. Therefore the increase of the potential energy of this portion of the shell in the given field is $-\mathrm{FI}\sigma t$, where t is the thickness of the shell. Now $\mathrm{I}t$, the strength of the shell, is equal to i, the strength of the current; and so the increase of the potential energy is $-\mathrm{F}i\sigma$. Consequently, the total change of the potential energy of the shell is

$$-i\mathrm{N},$$

where N is the whole force acting on the shell, that is, the number of lines of force which pass through the circuit whose action is represented by that of the shell.

363. *Induced Electromotive Force.*—Since, as we have just seen, the increase of the potential energy of an electric circuit through which a current i flows is $-i\nu$, where ν is the increase of the number of the lines of force which pass through the circuit in the positive direction, so, conversely, the work which is done upon the circuit by the electro-magnetic forces during a process in which the number of lines of

force which pass through the circuit increases by the amount ν is $i\nu$. If the conditions are such that this work can be transformed into electric energy in the circuit, a reverse electromotive force must be produced which opposes the passage of the current i. Now this electric energy is developed to the amount iE per unit of time, where E is the reverse electromotive force (§ 319). Therefore, since ν/τ is the increase of N per unit of time—if τ be the small interval of time in which the change ν takes place—we get

$$\frac{\nu}{\tau}=\mathrm{E}\,; \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; (1),$$

that is, *a reverse electromotive force acts around the circuit which at any instant is measured by the rate of increase of the number of lines of force which pass through the circuit.* [It must be remembered that a reverse electromotive force is one which tends to produce a current in the circuit in a direction which is related to the direction of the lines of force according to the law of left-handed screwing motion.]

The phenomenon, whose existence we have here assumed, was discovered experimentally by Faraday. He found that if, from whatever cause, the number of lines of force passing through a circuit is increased, a reverse electromotive force will act round the circuit, and will produce a reverse current; and that, if the number of lines be decreased, a direct force will act and will produce a direct current. The currents so produced are called *induced currents.* They only last so long as there is a variation of the number of lines of force in progress. In particular, they may be produced by the electromagnetic action which is due to varying currents in other fixed circuits, or to the motion of other circuits which carry steady currents; or they may be due to the action of moving magnets.

For example, if a current be started in one direction in a linear conductor, a transient current will flow in the opposite direction in a parallel conductor; and, if the direct current in the former be stopped, a transient direct current will flow in the latter. Phenomena such as these are called phenomena of *mutual induction.*

But it is important to observe that the number of lines of force which pass through a closed circuit depends upon the current which is flowing through that circuit as well as upon external currents. These lines of force pass in the positive direction through the circuit, and so any increase in their number, due to an increase in the strength of the current, causes a reverse electromotive force in the circuit which prevents the direct current from instantly attaining its full strength under the action of a suddenly introduced electromotive force, or from instantly falling to zero when the force is removed. This phenomenon is called *self-induction*, and was investigated experimentally by Faraday. It can only be prevented by arranging the circuit in such a way that its total area is zero. For example, if a plane circuit be crossed upon itself in a figure-of-eight shape, so that the areas of the two loops are equal, no self-induction will occur, for the lines of force which pass through each loop are equal in number, but are oppositely directed. The phenomenon of self-induction may be regarded as mutual induction between two parts of the single conductor placed side by side in the multiple arc arrangement.

The number of lines of force which pass through a circuit because of the magnetic action of the current flowing in it is

$$N = CS \quad . \; . \; . \; . \; . \; . \; . \; . \; (2),$$

where C is the value of the current and S is a quantity which depends upon the form of the circuit. This quantity is called the *coefficient of self-induction* of the circuit, or, more shortly, the *inductance* of the circuit.

Similarly the number of lines of force which traverse one circuit because of the electro-magnetic action of another circuit through which a current C is flowing is

$$N = CM, \quad . \; . \; . \; . \; . \; . \; . \; . \; (3),$$

where M is a quantity which depends upon the form and mutual position of the two circuits (§ 344), and is called the *co-efficient of mutual induction* of the two circuits.

Let a constant electromotive force E_1 act in the circuit through which the current C_1 is flowing, and let an electromotive force E_2 act in another circuit, through which a current C_2 flows, and whose coefficient of mutual induction with regard to the former circuit is M. Also let the resistance of the former circuit be R_1, while that of the latter is R_2; and let the co-efficient of self-induction of the former be S_1, while that of the latter is S_2. If M does not change, while C_1 changes by the amount c_1 in the small time τ, the corresponding change in the number of lines of force which pass through the second circuit is Mc_1; and, if C_1 does not change while M changes by the amount m, the corresponding change in the number of lines is C_1m. Also, if S_2 is constant while C_2 changes by the amount c_2, the change in the number of lines is S_2c_2; while, if C_2 be constant and S_2 changes by the amount s_2, the change in the number of lines is C_2s_2. Hence, by (1)

$$E_2 = \frac{Mc_1}{\tau} + \frac{C_1m}{\tau} + \frac{S_2c_2}{\tau} + \frac{C_2s_2}{\tau} + R_2C_2, \quad . \; . \; . \; . \; . \; (4),$$

and similarly

$$E_1=\frac{Mc_2}{\tau}+\frac{C_2m}{\tau}+\frac{S_1c_1}{\tau}+\frac{C_1s_1}{\tau}+R_1C_1 \quad \ldots\ldots (5).$$

If the two circuits are fixed, these equations become respectively

$$E_2=\frac{Mc_1}{\tau}+\frac{S_2c_2}{\tau}+R_2C_2 \quad \ldots\ldots\ldots (6)$$

$$E_1=\frac{Mc_2}{\tau}+\frac{S_1c_1}{\tau}+R_1C_1 \quad \ldots\ldots\ldots (7).$$

The first term on the right-hand side of each equation represents the reverse electromotive force due to mutual induction; the second represents the reverse force due to self-induction; and the third represents (§ 310) the part of the electromotive force which maintains the current against the resistance of the circuit.

364. *Laws of Rise and Decay of Currents.*—If the circuit in which E_2 acts be entirely removed, the equation which applies to the other circuit becomes

$$E=S\frac{c}{\tau}+RC.$$

Let us suppose that a constant electromotive force, E, is applied, the current being initially zero. The final value of the current-strength will, in accordance with Ohm's Law, be E/R. Until that stage is reached, the equation gives

$$\frac{c}{\tau}=\frac{R}{S}\left(\frac{E}{R}-C\right).$$

Now c/τ is the rate of increase of the current per unit of time; so the equation shows that the rate of increase is positive and is proportional to the defect of the actual current-strength from its final value. In other words, this defect decreases, as the time increases, at a rate which is proportional to its own value. The value of the current-strength therefore increases at a decreasing rate, and, theoretically, does not reach its final value in finite time. Practically, the final value is reached in most cases in a small fraction of a second.

Next let us suppose that the force E is maintained until the current attains its steady value C, after which E is withdrawn. We then have

$$S\frac{c}{\tau}+RC=0,$$

or

$$\frac{c}{\tau}=-\frac{R}{S}C.$$

This shows that the current decreases in strength at a rate which is proportional to its actual value. In fact the intensity of the current diminishes in geometrical progression as the time increases in arithmetical progression, and so it theoretically takes an infinite time to reach zero intensity. Practically, the condition of zero intensity is in most cases attained in a small fraction of a second.

Under both conditions the rate of variation increases in direct proportion to the resistance of the circuit, and decreases in direct proportion to the inductance of the circuit. The inductance therefore produces in an electric current the *appearance* of inertia.

365. *Special Cases of Induced Currents.*— The introduction of iron cores into the circuits greatly increases the self and mutual induction, because of the great permeability of iron. This is the essential principle of the *induction coil*, which consists of a coil of stout, insulated copper wire wound round an iron core, and surrounded by another coil of very fine, well-insulated copper wire. The inner coil is called the *primary* coil, the outer is called the *secondary* coil. The former has very small resistance, while the latter has very high resistance. The core is composed of a number of fine iron wires for the purpose of preventing the induction of currents within it—for these currents act so as to oppose the direct induction. By this means feeble electromotive forces in the primary circuit may give rise to very high electromotive forces in the secondary circuit. Other special cases will be considered subsequently.

Arago found that a magnet, which is pivoted above a horizontal copper disc, will be set into rotation if the disc be rotated. Faraday explained this by the electro-magnetic action of the currents which are induced in the disc. If radial slits be cut in the disc, the action will greatly cease; for the induction of currents is prevented except on a small scale.

As we have just seen, currents are induced in the body of a magnet itself whenever its state of magnetisation varies. Thus, since an electric current flowing along a rod is surrounded by closed lines of magnetic force, conversely, any change in the circular magnetisation of a rod will cause the flow of a transient current along the rod. This may readily be made manifest by connecting the ends of a twisted iron rod to the terminals of a galvanometer (§ 320), and suddenly magnetising the rod longitudinally. Since the rod is twisted, longitudinal magnetisation cannot occur without circular magnetisation (§ 358), and so a transient longitudinal current occurs, and is made manifest by the galvanometer. The same effect is produced if a longitudinally magnetised rod be suddenly twisted.

366. *Electrokinetic Energy.*— Consider again the two electric circuits dealt with in last section. Let them move so that M increases by the amount m, and let the motion be so slow that the currents C_1 and C_2 are sensibly constant. The rate at which heat is developed per unit of time in the two circuits is

$$H=R_1C_1^2+R_2C_2^2,$$

and the rate at which energy is supplied per unit of time in maintaining the electromotive forces E_1 and E_2 constant is

$$A=E_1C_1+E_2C_2.$$

Hence $$A-H=C_1(E_1-R_1C_1)+C_2(E_2-R_2C_2);$$

which becomes $$A-H=2C_1C_2\frac{m}{\tau},$$

since, in (4) and (5) above, if C_1, C_2, S_1, and S_2 are constant, we get

$$E_1-R_1C_1=C_2\frac{m}{\tau},$$

and $$E_2-R_2C_2=C_1\frac{m}{\tau}.$$

But by (1) and (3) we see that

$$C_1C_2\frac{m}{\tau}$$

represents the rate at which work is done in the circuit by electromagnetic action, and thus the equation shows that, under the given conditions, an amount of energy must be drawn from the source in a given time which exceeds that developed in the circuits in the form of heat by twice the amount of work which is simultaneously performed by the electro-magnetic forces. This excess is called the *electrokinetic* energy of the system. In Maxwell's theory it is supposed to reside in the medium which surrounds the circuits. It is transformed into heat, etc., whenever the circuits are broken; for the rupture of the circuits is, under these conditions, attended by a spark of more than usual intensity.

A similar result can be deduced from (6) and (7) when the circuits are fixed and C_1 and C_2 vary.

Again, we have seen (§ 362) that the energy of a circuit carrying a current C and traversed by N lines of force is changed by the amount NC. Consider the field N due to the current itself. The energy stored in the medium must be $\frac{1}{2}$NC, since N is not now constant, but increases in strict proportion to C. But, by (2), § 363, this is equal to $\frac{1}{2}SC^2$. This energy also is transformed into heat, etc., when the circuit is broken.

367. *The Ballistic Method.*—In the Ballistic Galvanometer the suspended portion has great moment of inertia, and, consequently, has a long period of vibration (§ 107). When a transient current, whose duration is very small in comparison with the periodic time, passes through this instrument, the total quantity of electricity which passes is proportional to the sine of half the angle of deflection. Hence the instrument may be applied to the investigation of magnetic properties. For example, if a coil of wire connected with the galvanometer be wound on an iron bar the intensity of magnetisation of which is varied from time to time, the transient current which follows each variation of intensity produces a deflection in terms of which the total quantity of electricity which passes can be calculated. Now, if ν be the change of induction through the coil in a very small interval of time τ, we get by (1), § 363,

$$\nu = \mathrm{E}\tau,$$

where E is the electromotive force, which is equal to RC—if C is the intensity of the current and R is the resistance of the circuit. The total change of induction is the sum of all the quantities of the type $\mathrm{RC}\tau$. If R is constant, this is the product of R into the sum of all the quantities of the type $\mathrm{C}\tau$. But $\mathrm{C}\tau$ is the quantity of electricity which flows through the circuit during the small time τ in which the current-strength had the value C. Hence the total magnetic induction is

$$\mathrm{N} = \mathrm{R}q,$$

where q is the total quantity of electricity which has passed. Its value may, therefore, be determined experimentally by means of the indications of the galvanometer. (It must be remembered that if the coil consist, for example, of n turns wound closely on the bar, the actual induction—supposed to be uniform—in the bar is $\mathrm{R}q/n$.) Hence (§§ 353, 359) we can determine the permeability and susceptibility of the substance of which the bar is composed.

368. *The Simple Harmonic Alternating Current.*—Suppose that we have a circular metallic circuit placed in a uniform field of magnetic force, and let the circuit be free to rotate about a diameter. Let this diameter be at right angles to the lines of force, and let the plane of the circuit be parallel to the lines. If now the circuit be made to revolve at a uniform rate about the given diameter, the number of lines passing through the circuit in one direction increases in accordance with the simple harmonic law; for the area of the circuit, taken perpendicular to the lines, varies in proportion the sine of the angle through which the plane of the circuit has

turned from its initial position. Now the electromotive force which is induced in the circuit is in proportion to the time rate at which the number of lines of force passing through the circuit varies; and this time rate varies in proportion to the cosine of the above angle.

To see this, consider a circle of unit radius, in which a very small chord is inscribed. Draw radii to the extremities of the chord. The lengths of perpendiculars, drawn from these extremities to a fixed diameter, represent the sines of the angles which the radii make with that diameter; and the projections of the radii on the diameter represent the cosines of these angles. But the ratio of the difference of the perpendiculars to the length of the small chord is practically equal to the cosine of either of the angles; and the ratio of that difference to the (constant) small interval of time in which it occurs is the time rate of variation of the sine. This proves the statement.

Hence the electromotive force varies, in the first quarter of a revolution, in proportion to the cosine of the angle through which the circuit has turned; and it is easy to see that it does so throughout a whole revolution.

Thus the induced electromotive force obeys the simple harmonic law; and, if Ohm's Law held good, the induced current would vary in exact proportion to the electromotive force. It changes sign at each odd quarter of a revolution, for then the number of lines crossing the circuit has its maximum value.

369. *Law of Conduction in an Inductive Circuit.*—It is evident that, in general, in a circuit in which inductive effects appear, the flow of electricity cannot obey Ohm's Law. We have seen, in § 364, that the effect of self induction is to prevent the current from attaining its full value at the instant at which it would otherwise do so, and to prevent it from becoming zero at the instant at which it would otherwise do so. In this way, when the direct electromotive force alternates, the reverse electromotive force shifts backwards the zero points and the maxima of the resultant electromotive force whose variations are strictly followed by the current. This retardation is termed *lag*.

Another obvious result of induction is a damping or choking effect. The current requires infinite time to rise to its maximum value, or to fall to zero, when there are no alternations. Consequently, as it is always opposed by the reverse electromotive force, it can never, when alternations occur, attain its proper maximum value.

We have already found that the reverse electromotive force is equal to the product of the inductance into the rate of variation of the current. Let us assume, therefore, that the actual current flowing in the circuit obeys the simple harmonic law. Let its value be $C_0 \sin 2\pi nt$, so that n is the number of complete periods which occur per second and C_0 is the maximum value of the current. Now, from the geometrical investigation given in last section, we see that the rate of variation of the sine of an angle is equal to the product of the cosine of that angle into the angular velocity with which the corresponding circle is described. Hence, if S be the inductance, the value of the reverse electromotive force is $2\pi n SC_0 \cos 2\pi nt$. The reverse force, therefore, differs in phase from the current by one quarter of a period.

To find the direct electromotive force which must be used to maintain the current $C_0 \sin pt$ $(p=2\pi n)$ in a circuit of resistance R, we have to find the value of the sum of $RC_0 \sin pt$ and $pSC_0 \cos pt$. The latter part is required to overcome the reverse electromotive force. Now, the instruments which are used for the measurement of alternating currents or electromotive forces only give (§ 380) the square-root-of-mean-square value of the strength of the quantity measured. As this average takes no account of difference of phase, the average value of $C_0 \sin pt$ is the same as that of $C_0 \cos pt$. Let this average be C, and we get, as the value of the instrumentally-indicated direct electromotive force, the quantity

$$E = C\sqrt{R^2+p^2S^2}.$$

Thus the resistance is apparently increased in the ratio of $\sqrt{R^2+p^2S^2}$ to R. The quantity under the radical sign is called the *impedance.*

Since the resultant of two simple harmonic quantities, which differ in phase by one quarter of a period, has a phase which is less than that of the more advanced component by the angle whose tangent is the ratio of the maximum magnitude of the less advanced component to that of the more advanced component, we can readily get an expression for the instantaneous current which flows under a given electromotive force in the given circuit. If this electromotive force be $E_0 \sin pt$, while C' is the instantaneous current, we get

$$E_0 \sin p(t-t_0) = C'\sqrt{R^2+p^2S^2},$$

where $\tan pt_0 = pS/R$.

If the inductance be so great that pS is large in comparison with R, the impedance practically depends on the inductance alone. Coils constructed on this principle—called *choking coils*—are used retard the phases of alternating currents.

370. *Law of Conduction in Condensing Circuits.*—If a condenser be placed in a conducting circuit, the reverse electromotive force tends to stop the current sooner than it would otherwise stop. The action is thus contrary to that which takes place in an inductive circuit, and so an acceleration of phase occurs.

The reverse electromotive force is due to the charging of the condenser as the current flows into it. If we suppose that one plate is kept at zero potential, the potential of the other plate is equal to the reverse electromotive force. - Let this potential be V, while K is the capacity of the condenser. The product VK is equal to the total charge of the second plate. Now, if $C_0 \sin 2\pi nt$ represents the current, the quantity of electricity which flows into this plate in a very small interval of time τ is $\tau C_0 \sin 2\pi nt$, and the total charge is the sum of all such quantities.

Now we found in § 368 that the rate of change of the sine of an angle, as the angle increases, is numerically equal to the cosine of that angle. Similarly, from the symmetrical relations of sines and cosines, we see that the rate of change of the cosine of an angle, as the angle increases, is numerically equal to the sine of the angle, but is negative. Hence the product of the sine of an angle into the small change of the angle is equal to the change of the cosine taken negatively. But the quantity $\tau C_0 \sin 2\pi nt$ may be written as $C_0/2\pi n \,.\, 2\pi nt \sin 2\pi nt$; and so the total charge is the product of $C_0/2\pi n$ into the sum of all the changes of $\cos 2\pi nt$ taken negatively. Thus, as the sum of all the changes of the cosine, from the time at which it was zero, gives the cosine itself, we see that the charge is $-C_0/2\pi n \,.\, \cos 2\pi nt$. Hence we have $VK = -C_0/2\pi n \,.\, \cos 2\pi nt$; and from this, using V and C to represent the instrumentally-recorded averages we get

$$V = -\frac{C}{pK}.$$

The reverse electromotive force differs in phase from the current by one quarter of a period. So we get as an expression for the direct electromotive force, the quantity

$$E = C\sqrt{R^2 + 1/p^2K^2}.$$

And the expression for the instantaneous value, C', of the current is given (*cf.* § 369) by

$$E_0 \sin p(t + t_0) = C'\sqrt{R^2 + 1/p^2K^2},$$

where $\tan pt_0 = -1/pKR$.

371. *Total Conduction in a Condensing and Inductive Circuit.* —In this case we have

$$E_0 \sin p(t - t_0) = C'\sqrt{R^2 + (pS - 1/pK)^2},$$

where $\tan pt_0 = (pS - 1/pK)/R$. Thus the impedance is zero when $KS = 1/p^2$. Under this condition the flow of electricity in the circuit obeys Ohm's Law.

We do not here touch upon the distribution of current density over a given section of the conducting circuit. When the inductance is large, or the period of alternation is sufficiently small, the current is practically confined to a thin layer at the external surface of the circuit.

372. *Alternate Current Transformers and Electromagnets.*—The energy which is transmitted, per unit of time, by a current of intensity C flowing under an electromotive force E is EC. Hence, when the energy is transmitted to great distances, the cost of the material of the conducting circuit renders it necessary to transmit a given amount of energy by means of a small current flowing under a high difference of potential rather than by means of a large current flowing under a small difference of potential; for the large current would give rise to great loss of energy through the development of heat unless a massive conductor were used. On the other hand, large currents and small difference of potential are required for many purposes, such as illumination. It is therefore necessary to have an instrument which will effect the required transformation.

The *transformer* which is employed consists essentially of an induction coil, whose core is a closed magnetic circuit of soft iron, whose primary coil is made of a large number of turns of fine, well-insulated, copper wire, and whose secondary coil is made of a small number of turns of stout copper wire or ribbon. The core must be longitudinally laminated in order to prevent waste of energy by means of currents induced in it.

Since the principal action of an electromagnet does not depend upon the direction of its magnetisation, it is possible to employ alternating currents in the coil of an electromagnet if the core be well laminated. With a given electromotive force, fewer turns of wire are used in a magnet designed for this purpose than in one designed for use with continuous currents. This is desirable because of the damping action of the arrangement.

Currents are induced in a sheet of copper if it be placed in front of the pole of an electromagnet which is actuated by an alternating current. And the phase of these currents is retarded by self-induction to such an extent that the sheet is repelled from the pole because of the magnetic effects of the currents.

373. *Alternate Current Dynamos.*—The 'dynamo' is a machine used for the transformation of mechanical work into electrical energy. The term 'alternator' is more usually employed when

alternating currents are produced. The simple circular ring, described in § 368, is an elementary form of alternator. If the ring were divided at one extremity of the diameter about which it rotates, and if one end of the wire were attached to an insulated metal ring fastened on the axis of rotation while the other end was attached to another similar ring fastened on the axis, the alternating current produced in the wire could be led away by means of metallic brushes pressing on the metal rings.

If, instead of a single circuit of wire, a coil of a number of turns be used, the effect will be increased in proportion to the total area. The coil does not, in any actual dynamo, revolve in a strictly uniform field; but the general nature of the action is sufficiently well indicated in the preceding discussion.

The magnets which are used for the production of the magnetic field are called the *field magnets;* and the coils, with their soft iron core (or cores), which traverse the field constitute the *armature.* The coils do not simply revolve in a field of force; they are carried across it—or rather, in general, across a series of strong fields in the course of a single revolution. The object of using multiple fields is to enable a sufficient number of alternations to take place at a comparatively slow speed of revolution. In other forms of alternators, to avoid the evil effects of strain in the armature on the insulation of the coils, the armature is fixed and the field magnets, which are sometimes replaced by a multiple-poled single field magnet, revolve.

The field magnets of alternators are excited by a separate continuous current.

374. *Continuous Current Dynamos.* — Dynamos giving continuous currents were employed long before alternating machines came into use. Any alternator could be made into a continuous current dynamo by providing it with a proper commutator which will convert the alternating currents in the armature coils into a direct current in the external circuit. In the case of the elementary arrangement considered in last section, for example, the commutation might be effected by connecting the ends of the divided wire to opposite sides of a single insulated ring, fixed on the axis of revolution, and divided into two halves along a diameter. If the brushes of the external circuit were so placed, pressing against the ring at opposite extremities of a diameter, that each slipped from one half of the ring to the other at the instant at which the current in the circuit became zero, the external current would be always direct. But, though direct, it would not be continuous in the sense of being steady. Twice in each revolution, it would fall to zero. Still, by

increasing the number of such circuits to two, four, eight, etc., symmetrically arranged, the current might be given as great steadiness as was necessary.

In most dynamos, a single field magnet with two poles is used, and the armature revolves in the space between the pole-pieces. In general the coils of the armature form a closed circuit. In some machines (*e.g.*, the Gramme) the coils are wound on an anchor-ring of iron, one diameter of which corresponds to the central line of the field. If the brushes make contact with the coil, as it revolves, at opposite extremities of the diameter which is perpendicular to the former, the induced currents, in both halves of the armature, act so as to send a continuous current through the external circuit. In actual machines the brushes do not make contact directly with the coil. They make contact with copper bars which are joined on to the coil at equal intervals.

In other dynamos, the open-coil system is used.

In the earlier machines, permanent magnets were used for the production of the magnetic field. These are called *magneto-electric* machines. The first great improvement lay in the introduction, by Siemens, of an iron core to the coils—which concentrated the field in their interior. Then Wilde employed electromagnets which were excited by a small magneto-electric machine. Lastly, the currents produced in the armature were used to excite the field magnets, the slight residual magnetisation of the latter being sufficient to start the action.

There are three special modes of winding the coils of the field magnets. In a *series-wound* dynamo, the whole of the current is sent round the field magnet coils, which are placed in series with the external circuit. In a *shunt-wound* dynamo, the coils are placed as a shunt to the main circuit, so that only a portion of the current is used to excite the magnets. In a *compound-wound* dynamo, the main current and also a shunted current flow round the coils. Sometimes a separate coil on the armature is used for the production of the exciting current.

375. *Characteristic Curves of Dynamos.*—The action of various types of dynamos is well shown by means of curves whose ordinates represent the electromotive force at the terminals while the abscissæ represent the value of the external current.

Consider, first, a series machine. When such a machine is driven at a constant speed, the electromotive force has a small finite value when the circuit is open. When the circuit is closed, if the resistance is gradually diminished in such a way that the current increases at a steady rate, the electromotive force increases. The increase is at first slow, then very rapid, then more slow. A maximum is reached when the current has a definite value, and, after that, the electromotive force available at the terminals slowly falls because of internal actions in the armature.

The tangent of the angle which a straight line drawn through the origin makes with the current-axis gives the resistance (ratio of electromotive force to current strength) of the external circuit. It therefore follows from the form of the characteristic curve that, in a series dynamo, the external resistance cannot be increased beyond a definite value without causing the current and electromotive force to drop to very small values.

In the case of a shunt-wound dynamo, the electromotive force has its highest value when the external circuit is unclosed; and, as the external resistance is diminished, the electromotive force gradually falls, for less current goes through the shunt coils.

376. *Constant Potential, and Constant Current, Dynamos.*—In incandescent lighting it is necessary that a constant difference of potential should be maintained at the terminals of the lamps. Hence it is necessary that the dynamo which supplies the current should maintain this constant difference of potential though the intensity of the current may vary through a large range. The forms of the characteristic curves for series and shunt dynamos show that a properly constructed compound-wound machine will satisfy the condition.

Dynamos with open-coil armatures (*e.g.*, the Brush machine) are used when, as in the case of lighting arc lamps placed in series, a constant current is required although more lamps may be turned on. The strength of the current is regulated automatically.

377. *Electro-dynamic Action on an Electric Circuit.*—When a force F produces a small change ϕ in one of the quantities which determine the position of the circuit, the work which is expended is $F\phi$, and the corresponding decrease in the potential energy of the circuit is $i\nu$, where ν is the increase of N. Hence

$$F = i\frac{\nu}{\phi}.$$

And we see that the force F tends to produce or to oppose the change ϕ according as that change is accompanied by an increase, or by a decrease, of the number of lines of force which pass through the circuit in the positive direction.

In particular, if a part of the circuit be moveable, the electromagnetic forces which act upon the circuit will tend to produce such a displacement of the moveable part as will cause an increase of the number of lines of force which pass through the circuit.

378. *Case of Linear Circuits.*—We have already seen that a linear circuit carrying an electric current is surrounded by circular lines of force, the direction of which is related to that of the current in the same way as the rotation of a right-handed screw is related to its linear motion.

Let A B (Fig. 187) represent part of a fixed linear circuit through which a current flows from A to B, and let *ab* be a portion of a moveable parallel circuit through which a current flows from *a* to *b*. We may assume that the circuit *ab* is completed by way of *p*. The lines of force due to AB pass through *abp* in the positive direction, and a displacement of *ab* towards AB would increase the area *abp*, and so would cause an increase in the number of positively drawn lines of force. Hence the electrodynamic action between the circuits is such as to cause their mutual approach. [We would arrive at the same result by the supposition that the circuit *ab* is completed by way of *q*. For the lines of force due to AB pass through the electric circuit *abq* in the negative direction. Hence *ab* will move so as to diminish the area of *abq*, that is, so as to diminish the number of negatively drawn lines of force which pass through it.] Similarly, mutual repulsion will ensue if the currents are oppositely directed.

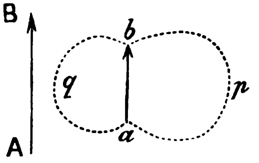


FIG. 187.

Next let the circuits AB and *ab* be inclined to each other, and let OO′ (Fig. 188) be the shortest line between them. Let us suppose that, while AB is fixed, *ab* is free to turn around OO′ as an axis. Complete the circuit *ab* by way of *p*. No lines of force due to AB will

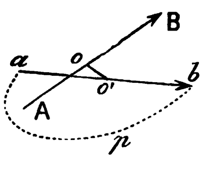


FIG. 188.

pass through *abp* when AB and *ab* are mutually perpendicular. On the other hand, the number of lines which pass through it in the positive direction is a maximum when the currents in AB and *ab* are parallel and similarly directed. The electrodynamic action on *ab* is therefore such as to cause it to place itself parallel to AB. If either current be reversed the moveable circuit will turn so that *ba* is co-directional with AB.

379. *Continuous Motion under Electro-dynamic Action.*—We have already seen (§ 346) that no work is, on the whole, done upon a magnetic pole which describes a closed path in a field of force due to an electric circuit, provided that the path does not pass through the interior of the circuit. But it is also true that no work will on the whole be done upon a magnetised body which completely describes a closed path passing through the interior of the circuit; for the body is composed of excessively small magnetised molecules, and the total amount of work which is expended upon each molecule in the process is zero, since its north and south poles are of equal strength.

But work will be expended on the whole if motion of part of the circuit takes place, without interruption of the current, under the action of external magnetic force. As an example, let us consider a horizontal circular conductor AB (Fig. 189). A current which enters this circuit at A, will divide into two parts, which reunite at B and flow through the conductor BC to the point C, which is connected with the negative pole of the battery to the positive pole of which the point A is joined. The lines of force, due to the earth's magnetic action, pass downwards through the circuit. In the region ABC, their direction is related to that of the current according to the law of left-handed screwing motion: in the region to the other side of BC, their direction is related to that of the current according to the law of right-handed screwing motion. Hence the

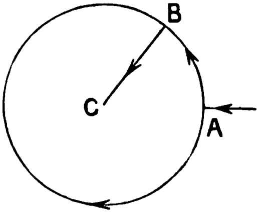


FIG. 189.

electrodynamic action upon BC will cause it to rotate in the direction of the hands of a watch provided that it is pivoted at C, and has a sliding contact at B.

If AB were a circular conducting disc, pivoted at C, and having a sliding contact at its circumference, so that an electric current flowed radially inwards, and if lines of force (whether due to the earth or to external magnets) passed through it as above, continuous rotation of the disc in the direction of the hands of a watch would ensue. This arrangement is known as Barlow's wheel.

Conversely, continuous motion of the magnet may take place if the circuit be fixed while the magnet is free to move and the direction of the current is reversed whenever the magnet passes from one side of the circuit to the other. Indeed, it is easy to arrange the conditions in such a way that continuous rotation of the magnet will take place without periodic reversal of the direction of the current. For example, let *ns* and *n′s′* (Fig. 190) be two magnets, which are connected by means of a cross-piece CD, and which, being pivoted at B, are free to rotate about AB; and let a current flow continuously along AB. The direction of the circular lines of force which surround AB is related to that of the current according to the law of right-handed screwing motion; and therefore the poles *n* and *n′* rotate round AB in that direction. And it follows that a delicately pivoted magnet, along one half of which a current flows, will be similarly set in rotation, for it may be supposed to consist of a number of magnets grouped around its axis.

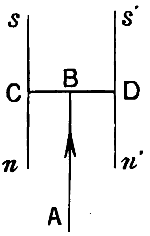


FIG. 190.

In the reverse action of the above, or similar, arrangements, electric currents are produced from mechanical work. This is the principle of the falsely-so-called *unipolar dynamos*. No great electromotive force is produced by these machines.

380. *Electric Motors.*—The principles of electrodynamic action are applied practically in the construction of electro-magnetic

machines or motors for the transformation of electric energy into mechanical work. The electric circuits may be fixed while the magnets rotate; or, preferably, the magnets may be fixed while the circuits rotate.

Every continuous current dynamo can be used as a motor, and can be driven by the current from another dynamo. The magnetic action of the current flowing in the armature of the motor sets the armature in rotation. This motion induces a reverse electromotive force in the coils of the armature, and it is the work which is done against this reverse force which appears in the form of mechanical work.

If the field magnets of the continuous current motor be laminated, an alternator may be employed, though not with advantage, to supply the current.

If two similar alternators are working in series, and are in the same phase, the action goes on steadily. But, if the phase of one gains on that of the other, the phase of the current in the resultant action, because of self-induction in the circuit, more nearly agrees with that of the electromotive force of the lagging machine than with that of the electromotive force of the leading machine. The result is that the leading machine is relatively opposing the current, and is driven as a motor by the lagging machine. And the difference of phase therefore increases until the two act in complete opposition. Thus an alternator is naturally driven as a motor by another alternator in series with it. On the other hand, if the two are coupled in parallel, the phase of the resultant current which circulates through the alternators because of a slight difference of phase from strict opposition, is, because of induction, more nearly in agreement with the phase of the electromotive force in the machine which is leading in phase. This machine therefore drives the other as a motor, and so the two machines are automatically kept running together.

One great defect of the alternating motor lies in the fact that the motor must have its speed independently raised to the proper value before it is thrown into circuit with the dynamo.

Alternating currents may be used to produce an alternating magnetic field. And, if two such fields, having the same period and intensity, but differing in phase by a quarter of a period, are arranged at right angles to each other, the result will (§ 34) be the production of a uniformly rotating magnetic field. The same thing will occur if three such fields, differing in phase by one-third of a period, be suitably arranged, and so on. This rotating field can be made to induce currents in a conductor, and so to set the conductor in motion, and motors are constructed on this principle. The motor can never rotate as fast as the field, for then the induction of currents would cease. For this reason such motors are termed *asynchronous.*

381. *Electrical Efficiency.*—If E be the total electromotive force generated in a dynamo, while C is the current-strength, EC is the total rate of working per second. But, if r be the internal resistance, the amount of work rC^2 is spent internally per second. Hence the efficiency, being the ratio of the external work to the total energy, is

$$1-\frac{rC}{E}.$$

In the case of a motor, to which energy EC is supplied per second, the energy absorbed per second is eC, where e is the reverse electromotive force generated in the motor. Hence the efficiency is e/E.

382. *Measurement of Alternating Currents and Electromotive Forces.*—The strength of an alternating current cannot be measured directly by the ordinary methods. The square of the current, or some quantity which does not change sign when the current changes sign, must be measured. The electrodynamometer (§ 320) is therefore directly applicable to the measurement of such currents. And, for the measurement of alternating electromotive forces, Cardew's voltmeter (whose indications, since they depend upon the elongation of a wire of high resistance by the heat developed by the passage of a current, are proportional to the square of the difference of potential at the extremities of the wire) may be used. A quadrant electrometer, if used on the ideostatic principle (§ 298), also gives indications which are proportional to the square of the potential difference.

In all cases, it is necessarily only an average—the square root of the mean square value—which is indicated. In the case of a simple harmonic variation, this average is 0·707 times the maximum value.

383. *Measurement of Power.*—Rate of working, or power, can be measured readily in the case of steady currents. It is only necessary to measure separately the electromotive force and the strength of the current. The product of these quantities gives the power.

In the case of alternating currents, the difference of phase between the current and the electromotive force makes the product of the separately measured averages exceed the actual power. An electrodynamometer with separate coils—one of small resistance to carry the current, and the other of high resistance to indicate electromotive force—is used to give the average value of the *simultaneous product.* It is termed a *Watt-meter* if it indicates power in watts (§ 386).

384. *Resistance expressed as a Velocity.*—Suppose that a uniform circular conducting circuit, of radius r, is placed in a uniform magnetic field of unit intensity, the plane of the circuit being perpendicular to the direction of the field. Let the radius of the circuit increase at a constant rate ρ per second. The electromotive force which is induced in the circuit, being equal to the rate at which the area of the circle increases, has the value $2\pi r\rho$; and the resistance, being proportional to the length of the circle, has the value $2\pi kr$, where k is the resistance of unit length of the circuit. Now let ρ have such a value that the current induced in the circuit is of unit intensity, and we get, by Ohm's Law, $k=\rho$. That is, a given resistance is numerically equal to a given velocity.

In the chapter on electrostatics we saw how a conductance could be expressed as a velocity, and yet a conductance is the reciprocal of a resistance. The reason for this apparent paradox will appear in the next section.

385. *Electric and Magnetic Units.*—The magnitude of electric and magnetic, as of all other quantities, depends upon the particular units in terms of which they are measured. All such quantities may be expressed in terms of the units of mass, length, and time; but the *dimensions* of a quantity in terms of these units depends upon the particular definition of some electric or magnetic quantity which we adopt.

Two systems of measurement are in use, the *Electrostatic* and the *Electromagnetic*. In the electrostatic system we start from the definition that two similar unit quantities of electricity, condensed at points which are at unit distance apart, repel each other (in air) with unit force (§ 280).

The dimensions of force are (§ 46) (MLT^{-2}), and, therefore, the dimensions of electric quantity are

$$(q)=(\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{3}{2}}\mathrm{T}^{-1}).$$

Surface density of electricity is quantity per unit surface. Its dimensions are, therefore, on this system,

$$(\sigma)=(\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{-\frac{1}{2}}\mathrm{T}^{-1}).$$

Electric potential and electric force have dimensions (§ 281)

$$(v)=(q\mathrm{L}^{-1})=(\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{1}{2}}\mathrm{T}^{-1})$$

and

$$(q\mathrm{L}^{-2})=(\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{-\frac{1}{2}}\mathrm{T}^{-1})$$

respectively.

The dimensions of electrostatic capacity are (§ 282)

$$(qv^{-1}) = (\mathrm{L}):$$

those of current strength are (§ 309)

$$(q\mathrm{T}^{-1}) = (\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{3}{2}}\mathrm{T}^{-2}):$$

and those of resistance are (§ 310) directly proportional to those of potential and inversely proportional to those of current-strength; they are, therefore,

$$(\mathrm{R}) = (\mathrm{L}^{-1}\mathrm{T}).$$

On the electromagnetic system the definition of unit quantity of magnetism is precisely analogous to the definition of unit quantity of electricity in the electrostatic system. Hence the dimensions of quantity of magnetism, surface density of magnetism, magnetic potential, and magnetic force, are, on this system, identical with the dimensions of the corresponding electric quantities on the electromagnetic system.

In addition, the dimensions of magnetic moment and intensity of magnetisation, on the latter system, are (§§ 343, 344)

$$(m) = (q\mathrm{L}) = (\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{5}{2}}\mathrm{T}^{-1})$$

and

$$(\mathfrak{I}) = (q\mathrm{L}^{-2}) = (\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{-\frac{1}{2}}\mathrm{T}^{-1})$$

respectively. The latter expression, of course, is identical with the expression for the dimensions of surface density.

On the electro-magnetic system, unit current is the current which, flowing in a circular circuit of unit radius, exerts, per unit length of its circumference, unit force upon a unit magnetic pole placed at its centre (§ 347). Hence the dimensions of current-strength are

$$(\mathrm{C}) = (\mathrm{MLT}^{-2}q^{-1}\mathrm{L}) = (\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{1}{2}}\mathrm{T}^{-1}).$$

The quantity of electricity which is conveyed through a conductor is directly proportional to the strength of the current and to the time during which it has been flowing. Therefore, its dimensions are

$$(\mathrm{Q}) = (\mathrm{CT}) = \mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{1}{2}}).$$

The dimensions of electric potential when multiplied by quantity of electricity are (§ 293) identical with those of energy, and are, therefore,

$$(\mathrm{V}) = (\mathrm{ML}^{2}\mathrm{T}^{-2}\mathrm{Q}^{-1}) = (\mathrm{M}^{\frac{1}{2}}\mathrm{L}^{\frac{3}{2}}\mathrm{T}^{-2}).$$

By such considerations we may readily determine the dimensions of any electrical or magnetic quantity on either system of reckoning. Some of the results are tabulated below, the dimensions on the electrostatic system being given in the second column, while those on the electro-magnetic system are given in the third.

Electrical Quantities.

Quantity of Electricity	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}})$
Surface density of Electricity	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$
Electric Potential	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2})$
Electric Force	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-2})$
Electrostatic Capacity	(L)	$(L^{-1}T^{2})$
Current Strength	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$
Resistance	$(L^{-1}T)$	(LT^{-1})
Specific Inductive Capacity...	$(M^{0}L^{0}T^{0})$	$(L^{-2}T^{2})$

Magnetic Quantities.

Quantity of Magnetism.........	$(M^{\frac{1}{2}}L^{\frac{1}{2}})$	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1})$.
Surface density of Magnetism	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Potential	$(M^{\frac{3}{2}}L^{\frac{3}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$
Magnetic Force.................	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Moment..............	$(M^{\frac{1}{2}}L^{\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{\frac{5}{2}}T^{-1})$
Intensity of Magnetisation ...	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Permeability.........	$(L^{-2}T^{2})$	$(M^{0}L^{0}T^{0})$
Magnetic Susceptibility	$(L^{-2}T^{2})$	$(M^{0}L^{0}T^{0})$

It is specially worthy of notice that the dimensions of any quantity on one or other of these systems always differ from its dimensions on the other by the dimensions of a speed or of a speed squared.

The force between two quantities, q and q', of electricity at a distance r apart in a medium other than air, is $qq'K/r^2$, where K is the specific inductive capacity. Hence, if we choose not to define the electrostatic dimensions of K as zero, but leave them undetermined, the electrostatic dimensions of quantity of electricity become

$$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}K^{\frac{1}{2}}).$$

Similarly, if we leave the dimensions of magnetic permeability (μ) undetermined, we find that the electro-magnetic dimensions of quantity of electricity are

$$(M^{\frac{1}{2}}L^{\frac{1}{2}}\mu^{-\frac{1}{2}});$$

and the corresponding alterations in the dimensions of other quantities can easily be found. One advantage of this method (due to Rücker) is, as Fitzgerald pointed out, that we can make the dimensions of any one quantity on both systems identical by assuming that the dimensions of K and μ are TL^{-1}. Also, as Rücker showed, by assuming that the reciprocal of the product $K\mu$ has the dimensions of the square of a speed, roots of the fundamental units disappear from the dimensions.

386. *Practical Units.*—It is convenient, in scientific measurements, to adopt the centimetre-gramme-second (c.g.s.) system of units; but, in practice, these units are often inconveniently large or inconveniently small. In the following table the names of the electro-magnetic practical units of various quantities are given in the second column; and the factors which are required to reduce the numerics, as expressed in terms of the practical units, to their equivalents on the c.g.s. system, are given in the third column.

Quantity of Electricity	Coulomb	10^{-1}
Electromotive Force	Volt	10^{8}
Electrostatic Capacity	Farad	10^{-9}
	Microfarad	10^{-15}
Current Strength	Ampere	10^{-1}
Resistance	Ohm	10^{9}
Magnetic Field	Gauss	10^{0}
Magnetic Induction	Henry	10^{9}
Electrical Energy	Joule	10^{7}
Electrical Power	Watt	10

An electromotive force of one volt maintains a current, whose strength is one ampere, through a resistance of one ohm.

The legal ohm is defined as the resistance of a uniform column of mercury, at 0° C., whose mass is 14·4521 grammes, and whose length is 106·3 cm.

The ampere is practically the strength of a steady current which, flowing through a solution of nitrate of silver, deposits 0·001118 gramme of silver per second; and the coulomb is the quantity of electricity which passes per second in this process.

The farad is the capacity of a condenser which is charged to the potential of one volt by one coulomb.

The joule is the amount of energy which, per second, is transformed into heat in a conductor whose resistance is one ohm and through which a current of one ampere flows.

The watt is such that 746 watts are practically equal to one horse-power.

The henry is the amount of induction which gives rise to an electromotive force of one volt when the inducing current varies at the rate of one ampere per second.

When the absolute values of two of the quantities involved in the expression for Ohm's Law have been determined, the absolute value of the third follows from its definition; and it is theoretically of no consequence which of the three pairs is chosen for determination experimentally.

Since the method described in § 359 enables us to measure easily, in absolute units, the strength of a magnetic field, we could, by measuring the strength of the field inside a solenoid determine (§ 347) the absolute value of the steady current flowing in the solenoid. And we could then obtain a ready practical means of measurement by making use of Faraday's law of electrolysis.

In order to obtain a practical unit of resistance we only require to find the value of a given resistance, expressed (say) as a velocity, in centimetres per second. A simple method of doing this consists in measuring, in a given conductor, the heat developed per second by a current of given intensity, and calculating the resistance by Joule's Law. In a direct method, due to Lord Kelvin, a closed coil is spun, in the earth's field, with its plane vertical and in the magnetic meridian. The induced current deflects a horizontal magnet suspended at the centre of the coil, and its magnetic effect is determined by the amount of the deflection. This effect is directly proportional to the horizontal intensity of the earth's field, to the number of turns in the coil, and to the speed of revolution, and is inversely proportional to the resistance. Thus, the resistance can be expressed in terms of the speed. The ohm is then defined to be 10^9 centimetres per second—a velocity intended to represent an earth's quadrant per second. The first British Association (B.A.) standard ohms which were constructed had a value of 0·9863 true ohm; and the Siemens unit is 0·9408 true ohm. Recent determinations give over 106·2 cm. of mercury, one square mm. in section, at 0° C., as a conductor whose resistance is one ohm. As already stated, 106·3 cm. is now defined as the true length of the mercury column, though it is presumably slightly too large.

387. *Explanation of the Phenomena of Electric and Magnetic Fields by the Motion of Faraday Tubes.*—We have already seen (§§ 286-292) how some phenomena can be explained, or rather co-ordinated, by means of the idea of tubes of electrostatic induction. By the tendency of such tubes to contract, and by their mutual repulsion, the known mechanical forces between electrified bodies can be accounted for. They are supposed to exist in the ether, and to electrically polarise matter through which they pass. Faraday also introduced the idea of tubes of magnetic induction, and by their motion also, the phenomena of the electro-magnetic field may be accounted for. J. J. Thomson has worked out the theory of the field upon the assumption of the existence of electrostatic tubes.

In this theory, the existence of closed tubes, as well as of finite open tubes, is assumed. The open tubes must end upon matter, and a unit tube corresponds to the atomic charge of a monovalent element. When the length of the tube joining two atoms is large in comparison with atomic distances in molecules, the atoms are chemically free. The tubes are regarded as incapable of creation or of destruction. Electromotive intensity is measured by the excess of the number of tubes which point in the positive direction over those which point in the negative direction.

Ether is assumed to move in and around the tubes even when they are at rest. The kinetic energy of this ether motion is the potential energy of the electrostatic field. The kinetic energy of the moving tubes is the potential energy of the magnetic field.

The motion of a tube causes a magnetic force whose direction is perpendicular to the directions of the tube and the motion. And, when the strength of the field varies, an electromotive force acts at right angles to the direction of motion and to that of the magnetic field, and gives rise to the known electro-magnetic induction of currents. Thus the tubes always tend to set their length perpendicular to the direction of motion.

If there is no free electricity in an isotropic field, so that the electromotive intensity is due to the motion of the tubes alone, the tubes move at right angles to themselves with a speed whose square is the reciprocal of the product of the permeability and the specific inductive capacity of the medium—the length of the tubes, the direction of motion, and the magnetic force being at right angles to each other. On the electro-magnetic theory this speed is that of light. Hence, as J. J. Thomson points out, the tubes take the place of corpuscles in the corpuscular theory of light, and can account for polarisation, since they possess 'sides.'

A steady magnetic field is produced when positively directed, and negatively directed, tubes move past each other at equal rates.

The communication of the momentum of the tubes to conductors accounts for the ponderomotive forces which act on these.

A tube shrinks to molecular dimensions when it reaches a conductor, and gives rise to a current in the process. Resistance can be expressed in terms of the time which a tube requires to shrink.

The phenomena of voltaic action, electrolysis, discharge of jars, etc., can readily be accounted for.

388. *Historical Development of the Subjects.*—The order in which we have considered the phenomena of electro-magnetism and electro-dynamics is not that in which the subjects were historically developed. Oersted found that a magnetic needle always tends to place its length in a direction at right angles to a plane which, passing through its centre, contains a linear circuit, through which an electric current is flowing. The direction in which the north pole points depends upon the direction in which the current flows. The north pole always tends to move round the linear circuit in a direction which is related to that of the current in the way in which the rotation of a right-handed screw is related to its linear motion.

After this fact was discovered, it was surmised that the converse phenomenon might also be found to exist—that motion of the linear circuit, through which the current flows, would take place if the magnet were fixed while the circuit was free to move. This was verified experimentally by Ampère.

And, further, since a magnet can act thus upon two neighbouring circuits, through which electric currents flow, it was supposed that mutual action might be found to exist between these circuits themselves if the magnet were removed. Ampère proved the existence of this action also.

CHAPTER XXX.

ELECTRO-MAGNETIC THEORY OF LIGHT.

889. *Magnetic Rotation of the Plane of Polarisation of Light.*—Faraday made many attempts to detect some action upon polarised light when it was made to pass through a dielectric which was subjected to electric stress. He also sought for evidence of such action when polarised light passed through an electrolyte conveying a current, but in no case could he observe any effect. On the other hand, he found a marked effect when polarised light was passed through a diamagnetic medium placed in a field of magnetic force.

When the direction of the ray coincides with the positive direction of the lines of force, the plane of polarisation is rotated through an angle which is proportional to the intensity of the magnetic field, and to the length of the path of the ray within the medium. If the direction of the ray does not coincide with the direction of the field, the rotation is proportional to the intensity of the resolved part of the force taken in the direction of the ray. The amount of the rotation per unit of length, in a field of unit intensity, depends upon the nature of the medium. The absolute direction of rotation is unaltered by a reversal of the ray, provided that the direction of the field is unaltered.

In diamagnetic media the direction of the rotation is, in general, connected with that of the field according to the law of right-handed screwing motion; in paramagnetic media, the reverse is generally true.

The fact of the non-reversal of the rotation with reference to the direction of the field points to a fundamental distinction between the mechanical method by which this magnetic rotation is produced and that obtained in cases of rotation by quartz or solutions such as sugar (§ 225). In the latter cases, reversal of the ray is not accompanied by a reversal of the rotation with reference to *it;* and thus the total rotation during the double passage through the medium is zero. The total magnetic rotation is doubled by the double passage.

In the case of vapours and gases, the rotation is very small in comparison with the rotation which is produced, under similar conditions, by solids and liquids. Even in vapours of liquids which have considerable rotatory power, the effect is very small.

Verdet has shown that the rotation is, approximately, inversely proportional to the square of the wave-length—the deviation being in *defect* as the wave-length increases, and being most marked in substances of great dispersive power.

390. *Hypothesis of Molecular Vortices.*—We have seen that a plane polarised ray may be compounded of two uniform equiperiodic circular motions of equal amplitude, and that rotation of the plane of polarisation will take place if one of these component motions is accelerated relatively to the other (§ 225). The explanation of the magnetic effect seems to lie in this direction.

Lord Kelvin has remarked on this subject 'That the magnetic influence on light discovered by Faraday depends on the direction of motion of moving particles. For instance, in a medium possessing it, particles in a straight line parallel to the lines of force, displaced to a helix round this line as axis, and then projected tangentially with such velocities as to describe circles, will have different motions according as their motions are round in one direction (the same as the nominal direction of the galvanic current in the magnetising coil), or in the contrary direction. But the elastic reaction of the medium must be the same for the same displacements whatever be the velocities and directions of the particles; that is to say, the forces which are balanced by centrifugal force of the circular motions are equal, while the luminiferous motions are unequal. The absolute circular motions being therefore either equal or such as to transmit equal centrifugal forces to the particles initially considered, it follows that the luminiferous motions are only components of the whole motion, and that a less luminiferous component in one direction, compounded with a motion existing in the medium when transmitting no light, gives an equal resultant to that of a greater luminiferous motion in the contrary direction compounded with the same non-luminous motion. I think it not only impossible to conceive any other than this dynamical explanation of the fact that circularly-polarised light transmitted through magnetised glass parallel to the lines of magnetising force, with the same quality, right-handed always, or left-handed always, is propagated at different rates according as its course is in the direction, or is contrary to the direction in which a north magnetic pole is drawn; but I believe it can be demonstrated that no other explanation of that fact is possible. Hence it appears that Faraday's optical discovery affords a demonstration of the reality of Ampère's explanation of the ultimate nature of magnetism; and gives a definition of magnetisation in the dynamical theory of heat. The introduction of the principle of moments of momenta ("the conservation of areas") into the mechanical treatment of Mr. Rankine's hypothesis of "molecular vortices" (see § 228), appears to indicate a line perpendicular to the plane of resultant rotatory momentum ("the invariable plane") of the thermal motions as the magnetic axis of a magnetised body, and suggests the resultant moment of momenta of these motions as the definite measure of the "magnetic moment." The explanation of all phenomena of electro-magnetic attraction and repulsion, and of electro-magnetic induction, is to be looked for simply in the inertia and pressure of the matter of which the motions constitute heat. Whether this matter is or is not electricity, whether it is a continuous fluid interpermeating the spaces between molecular nuclei, or is itself molecularly grouped; or whether all matter is continuous, and molecular heterogeneousness consists in finite vortical or other relative motions of contiguous parts of a body, it is impossible to decide, and perhaps in vain to speculate, in the present state of science.'

The idea contained in these remarks has been developed by Maxwell into a complete theory of molecular vortices. He points out that, from the fact that the wave-length, λ, and the periodic time, τ, increase and decrease together, it follows that if for a given numerical value of the angular velocity, n $(=2\pi/\tau)$, the value of the speed of propagation, λ/τ, is greater when n is positive than when it is negative, for a given value of λ the positive value of n will be greater than the negative value. This is so since the former condition implies that λ is greater when n is positive than when it is negative, and a diminution of λ implies a diminution of τ and therefore an increase of n. Since the ray does not diminish in intensity as it passes through the medium, the amplitude, r, must remain constant (*cf.* § 136). And the principles of dynamics show that, in this case, the increase of kinetic energy which is produced by a slight increase in the value of r, all other quantities being constant, is algebraically equal to the increase of potential energy thereby produced. In fact this statement is simply equivalent to the assertion that, under the given conditions, centrifugal force is balanced by central attraction; and it is easy, for example, to prove, by means of it, Kepler's Third Law in the case of circular planetary orbits. Now the expression for the kinetic energy contains one term involving n^2, and it may contain terms involving the products of n into other velocities, and terms independent of n. The potential energy, on the other hand, is independent of n. Hence, as we have just found that the difference between the kinetic and potential energies is not altered by a mere change of r, we see that n must satisfy a condition of the form

$$An^2+Bn+C=o,$$

where A, B, and C, are functions of the co-ordinates. Now experiment shows that n has two real values, one positive, the other negative and smaller. C must therefore be finite, and both it and B must be negative if A is positive; for $-B/A$ and C/A are respectively the sum and the product of the roots of the equation, and the sum is positive while the product is negative. B also cannot vanish, since the roots are distinct. The term in n must therefore involve another velocity besides n, and that velocity must be an angular velocity about the same axis, for Bn is a scalar quantity.

Maxwell then concludes 'That in the medium, when under the action of magnetic force, some rotatory motion is going on, the axis of rotation being in the direction of the magnetic forces; and that the rate of propagation of circularly polarised light, when the direction of its vibratory rotation and the direction of the magnetic rotation of the medium are the same, is different from the rate of propagation when these directions are opposite.

'The only resemblance which we can trace between a medium through which circularly-polarised light is propagated, and a medium through which lines of magnetic force pass, is that in both there is a motion of rotation about an axis. But here the resemblance stops, for the rotation in the optical phenomenon is that of the vector which represents the disturbance. This vector is always perpendicular to the direction of the ray, and rotates about it a known number of times in a second. In the magnetic phenomenon, that which rotates has no properties by which its sides can be distinguished, so that we cannot determine how many times it rotates in a second.

'There is nothing, therefore, in the magnetic phenomenon which corresponds to the wave-length and the wave-propagation in the optical phenomenon. A medium in which a constant magnetic force is acting is not, in consequence of that force, filled with waves travelling in one direction, as when light is propagated through it. The only resemblance between the optical and the magnetic phenomenon is, that at each point of the medium something exists of the nature of an angular velocity about an axis in the direction of the magnetic force.'

'This angular velocity cannot be that of any portion of the medium of sensible dimensions rotating as a whole. We must therefore conceive the rotation to be that of very small portions of the medium, each rotating on its own axis. This is the hypothesis of molecular vortices.

'The motion of these vortices, though . . . it does not sensibly affect the visible motions of large bodies, may be such as to affect that vibratory motion on which the propagation of light, according to the undulatory theory, depends. The displacements of the medium, during the propagation of light, will produce a disturbance of the vortices, and the vortices when so disturbed may react on the medium so as to affect the mode of propagation of the ray.'

From this hypothesis Maxwell has deduced an expression for the magnitude of the rotation under given conditions which accords very well with the results of observation.

391. *Hall's Effect.*—Hall found by experiment that a thin metallic conductor, which is placed in a magnetic field of force with its plane perpendicular to the lines of force, and through which an electric current flows, is the seat of an electromotive force which acts along the common perpendicular to the directions of the current and the magnetic field. Rowland has proved that, if a similar electromotive force appears when the electric displacement in an insulating medium varies in a field of magnetic force, rotation of the direction of the displacement will follow the passage of a wave through the medium; and Glazebrook has shown that the electromotive force is a consequence of the molecular rotation which Maxwell assumes.

392. *Kerr's Effects.*—The action upon polarised light in a medium subjected to electric stress, for which Faraday sought in vain, was discovered by Kerr. He placed two parallel brass plates at a short distance apart, in a glass cell containing carbon bisulphide, and connected them with the poles of an electric machine. A beam of light, polarised at an angle of 45° to the direction of the lines of electric force, was passed between the plates, and, on emergence, was found to be elliptically polarised. The axes were respectively parallel to and perpendicular to the lines of force; and the difference between the phases of the two components was proportional to the square of the intensity of the electric field.

Another form of this experiment consists in passing the polarised beam between two small spheres which are placed near to each other in the insulating medium, and are connected to the poles of an electric machine. The medium is then found to have become doubly refractive.

Dr. Kerr has also discovered that the plane of polarisation of light is rotated in the act of reflection from the (highly polished) pole of an electromagnet—the direction of the rotation being reversed when the magnetisation of the pole is reversed.

393. *Photo-electric Actions.*—The electric resistance of annealed selenium is greatly reduced by exposure to light. Tellurium and carbon are similarly affected to some extent.

Becquerel showed that chloride of silver, freshly deposited on a silver plate, is sensitive to light. If two such plates are placed in water, and are connected externally by a conductor, an electric current flows through the water from the illuminated plate to the other plate. Minchin has recently constructed other photo-chemical cells, some of which are sensitive to electro-magnetic undulations (§ 395) of greater wave-length than those which affect the retina.

Another photo-electric action has been described in § 324.

394. *Electro-magnetic Theory of Light.*—The phenomena which are described in the immediately preceding sections point to a very close connection between electricity, magnetism, and light.

We have already seen that the phenomena of light are best explained on the assumption that light consists in undulations propagated through a medium. On the other hand, the theories of electrical and magnetic action were originally expressed in terms of direct action at a distance; and (although Faraday conducted all his reasoning on the assumption of action through a medium) it was not until Maxwell translated Faraday's ideas into mathematical language that it was recognised that electrical and magnetic phenomena could be readily explained as the results of the propagation of action through a medium.

In determining the conditions of the propagation of an electromagnetic disturbance through the medium whose existence he postulated, Maxwell arrived at the conclusion that the propagation takes place in accordance with the laws of the transference of motion through an elastic solid, and that the speed of propagation is

$$V = \frac{1}{\sqrt{K\mu}},$$

where K and μ are respectively the specific inductive capacity and the permeability of the medium. In the propagation of a plane wave, electric displacement takes place at right angles to the direction of magnetic induction, and both are in the plane of the wave.

A reference to § 385 will make it evident that, on the electrostatic and electro-magnetic systems respectively, the dimensions of μ and K are those of the inverse square of a speed. According to Maxwell, if light is an electro-magnetic phenomenon, V must represent the speed of light. Now the speed of light, v, is capable of measurement to a considerable degree of accuracy, while the value of V can be determined directly by a comparison of the relative values of some electrical or magnetic quantity on the two systems of measurement; and the results of various independent determinations of the values of V and v strongly confirm the supposition of their numerical identity in air.

In media other than air the speed of light is inversely proportional to the refractive indices. Hence K should be practically equal to the square of the refractive index, for the value of μ is nearly unity in all transparent media. Since the experimental determination of K occupies a time which is practically infinite in comparison with the period of any luminous vibration, we must, in testing this point, take the refractive index for rays of infinite wave-length. (This might be given by the value of the constant a in Cauchy's expression for the refractive index, § 183.) Hopkinson has found that while the relation holds in the case of hydrocarbons, it does not obtain in glass and the animal and vegetable oils. In the latter substances the refractive index is less than $\sqrt{K}$.

In the luminiferous medium, two forms of energy exist—one kinetic, the other potential. Similarly, in the electro-magnetic medium, energy exists in a kinetic (electrokinetic, § 366) form and in a potential (electrostatic, § 293) form.

The theory also explains double refraction, and leads to Fresnel's construction for the wave surface. And it shows that the speed of propagation of a condensational-rarefactional wave would be infinite; so that this wave does not exist—a result which is in harmony with optical observations. Also, if the medium be not a perfect conductor, the electrical energy is in part transformed into energy of electric currents, and so finally into heat. This explains the absorption of light.

395. *Electro-magnetic Waves.*—The above evidence in favour of the truth of Maxwell's electro-magnetic theory is very strong in itself. But recent investigations, by Hertz and others, have proved, beyond the possibility of doubt, that electro-magnetic action is propagated with finite speed through a medium, and have indicated that its speed of propagation is identical with that of light.

If the initial disturbance is periodic, a series of electro-magnetic waves are propagated outwards from the source. The condition of periodicity can be obtained by means of the disruptive discharge, which, under suitable conditions, is oscillatory in its nature (§ 294), and has a constant period of oscillation depending on the electrostatic capacity and the coefficient of self-induction of the apparatus in which the discharge oscillates.

If the oscillator is a rectilinear wire, provided at each end with a ball of small capacity, the length of the wave which is emitted is double the length of the wire, but the disturbances die out before many oscillations have taken place. If the oscillator is a sphere, the wave-length is slightly more than three and a half times the diameter, and only one oscillation can be distinctly observed. If the spark passes, between two spheres, across the diameter of a hollow cylinder, the wave-length is approximately equal to the diameter of the cylinder, and a large number of oscillations can be observed.

Let us suppose, for the sake of definiteness, that the discharge takes place across the gap of a rectilinear oscillator. The alternating currents which characterise the discharge will induce similar currents in neighbouring conductors. According to the theory of direct action at a distance, these induced currents will appear in exact simultaneity with the inducing currents; according to the electro-magnetic theory, they will appear later and later in proportion as the conductors in which they are induced are more and more remote from the oscillator.

In his investigations on this point, Hertz made use of the principle of resonance (*cf.* § 148). That is, he used a secondary conducting circuit, the natural period of electric oscillation in which was the same as that in the primary circuit. The result is that the magnitude of the induced oscillations may be made very great, for each succeeding induced oscillation is timed to re-enforce the effects of preceding oscillations. In this way sufficient electromotive force may be developed to cause the electricity to spark across a small air-gap in the circuit.

Such a circuit, with its air-gap, may be used to make evident the existence of inductive effects at any given point in space, provided that its distance from the source is not too great. A vacuum tube acts very well if it be in circuit with a battery which is just unable to make the current pass of itself. An incandescent lamp whose fibre has been broken across works well in the same way. Electrometers and galvanometers, or the Cardew voltmeter, may be used. The most delicate arrangement consists of a tube, filled with iron filings, which is placed in the circuit along with a voltaic cell and a galvanometer. The current which flows, if at all appreciable, is very feeble until a spark passes in the neighbourhood, *i.e.*, until an electric wave reaches the circuit, when the current at once becomes large. The induced current breaks down the insulation between the filings and allows the voltaic current to flow, until, by tapping the tube, the observer readjusts the filings and makes the resistance great. The filings must be carefully screened by conductors from the effects of all electrical oscillations except those under observation.

If one source of oscillations alone existed in space, the intensity of the inductive effect at any given point, and therefore the intensity of the spark in the secondary circuit, would diminish continuously as the distance between the point and the source increased. If two sources existed, the intensity might be great in the neighbourhood of each, but would reach a minimum at some intermediate position: or, if the effects of the sources were opposite, the intensity of the resultant effect might be zero at some point between the two. If the effects were instantaneously propagated, only one such minimum could exist. But, if the effects were propagated by wave-motion at a finite rate, a great number of maxima and minima might appear, in accordance with the ordinary laws of interference. The best results will be obtained when the two sources are precisely similar. Now, if a conducting sheet be placed in the neighbourhood of a single source, the currents which are induced in it will, in turn, give rise to electro-magnetic effects having a periodic time equal to that of the source. This, on the electro-magnetic theory, constitutes reflection of the electro-magnetic radiation, and interference may be expected to take place between the incident and the reflected waves —nodes and loops occurring alternately at equal intervals of one-half of the length of a wave (§ 35). In performing this experiment Hertz was able to observe the existence of successive maxima and minima, and so the existence of electro-magnetic radiation was proved.

If the original electrical oscillations take place along a straight rod, the oscillations in the electro-magnetic medium will be parallel to the axis of the rod; *i.e.*, the wave is plane polarised. And the rod is surrounded by circular lines of magnetic induction, the direction of the induction changing with each alternation of electric displacement. The electric displacement and the magnetic induction therefore take place in the front of the wave, and are at right angles to each other; and these two effects can be separated from each other by using a suitable resonating circuit. A circular circuit held opposite the centre of the rod, with its plane passing through the axis of the rod, and with the line of its spark-gap at right angles to the axis, will respond only to the magnetic variations.

On the other hand, if held in front of the axis with the line of its spark-gap and one of its diameters parallel to the axis, it will respond only to the electrical variations, for no lines of magnetic induction pass through it, and the system is free to respond to the electrical impulse applied at the point of the circuit opposite the gap. If held in front of the axis, with one diameter parallel to the axis, and with the line of its spark-gap perpendicular to the axis, it will not respond at all, for no lines of magnetic induction pass through it, and the electrical impulses act symmetrically on either side of the spark-gap.

With apparatus such as is ordinarily used in a laboratory, radiations having a wave-length varying from a few inches to a number of miles in length, can be readily obtained. These long-period radiations can pass freely through insulators, such as pitch, which are opaque to luminous radiations. And, by using large prisms of such substances, their refractive indices can be found by the usual methods.

If the radiation falls upon a transparent sheet of a thickness which is small in comparison with the wave-length, no reflection is observed; for the acceleration of the phase by half a-period, which takes place at the second surface, produces total interference. This is a reproduction of the phenomenon of the central black spot in Newton's rings (§ 195).

Reflection can be obtained at the surface of a thick insulator if the direction of the electric displacements is perpendicular to the plane of reflection; but none is found at the polarising angle if the line of displacement lies in the plane of reflection. This proves that the electric displacement takes place at right angles to the plane of polarisation, and settles the much-debated question of the direction of the luminous vibrations in favour of Fresnel's assumption (§ 213).

Effects of diffraction can also be observed with these waves, and are in strict accordance with the results of the undulatory theory.

396. *Waves in Wires.*—It is easy to show the existence of waves in wires by attaching a long wire to a conductor which is placed near the conductor forming one end of a rectilinear oscillator. The conductor to which the wire is attached is charged by induction oppositely to the other conductor, and its periodic charges flow outward along the wire. When the wave reaches the end of the wire, reflection takes place, and, if the length of the wire be a multiple of half a wave-length, stationary waves are formed by the interference of the direct and the reverse waves. The circular resonator, above alluded to, can be used to find the positions of the

nodes of the electric and of the magnetic components of the waves. The magnetic nodes are at points where the circuit does not respond when held with its plane passing through the wire and the line of its spark-gap parallel to the wire. The electric nodes are at points where the circuit does not respond when held with its plane perpendicular to the wire and the line of its spark-gap also perpendicular to the wire. The electric action is greatest at the magnetic nodes, and the magnetic action is greatest at the electric nodes.

The direct wave propagated from the oscillator through the ether also affects the resonating circle, except when the plane of the circle is not, on the whole, crossed by any lines of magnetic force due to the oscillator, and is either perpendicular to the lines of electric force or parallel to them with the line of the spark-gap perpendicular to them. The nature of the interference is such as to show that the wave which is propagated by means of the wire travels at the same rate as the wave which goes directly.

The fact of the finite rate of propagation of electric waves along a wire is also shown readily by a method used by Von Bezold and, independently, by Hertz. If a symmetrical plane circuit of wire has its ends furnished with well-polished brass balls having a small spark-gap between them, and if the middle point of the circuit be joined by a wire to one pole of an induction coil in action, no spark will pass across the gap. But, if a different point be joined to the pole, sparks will pass at once. The reason is that one wave reaches the one knob by the short path sooner than the other wave, which goes by the long path, reaches the other knob. In this way a difference of potential sufficient to cause a spark is established.

This apparatus can be used to show that, in the case of such rapid oscillations, resistance has practically no effect in determining the speed of propagation. The determining causes are self-induction and capacity. If, on one side of the apparatus, the wire is arranged so as to have little self-induction, while, on the other side, it is arranged so as to have great self-induction, the null-point is no longer at the middle of the wire, but is shifted towards the side having large induction. In the same way, change of capacity is found to shift the null-point. If the two sides are similar in every way except that one consists of iron, while the other is made of copper, no displacement of the null-point occurs. This shows that the period of oscillation is so short that the magnetic changes in the iron cannot follow the rapid changes of magnetic force.

We have spoken of a wave as travelling along a wire. According to modern views, the energy does not flow along the wire, but through the surrounding medium. Poynting has shown that, on

Maxwell's theory, the flow of energy takes place along the lines of intersection of the electric and magnetic surfaces of equal potential. The wire acts merely as a guide for the energy. In so far as the wire possesses resistance, the lines of flow enter the wire from the side, and some of the energy is transformed into heat. Induction also causes an inward flow which gives rise to the energy of the induced current.

CHAPTER XXXI.

RELATED PHYSICAL PROPERTIES.

397. THE investigation of the relations which subsist amongst various physical quantities, and which are exemplified in the different transformations of energy, is one of the most interesting investigations in the whole range of physics. On the other hand, the mathematical methods which are usually employed for this purpose are of such a nature as to be unavailable to any who are unacquainted with higher mathematical methods. Yet many of these relations may be worked out by means of the most elementary methods.

The first relations of this kind which were investigated were those exemplified in the transformation of heat into mechanical work. In any such transformation, the dynamical principle of conservation of energy holds good; but it was found necessary to introduce another principle—the Second Law of Thermodynamics—before the problem could be fully solved; and this principle is not explicitly dynamical, but is based upon an axiom in agreement with experience. Further, when we pass from thermo-mechanical transformations to transformations between two forms of energy of which heat is not one, no simple law corresponding to the Second Law of Thermodynamics has been formulated. On the contrary, it is sometimes stated that, in these cases, this method reduces to the use of the principle of conservation alone—a principle which is often not sufficient for the purpose. We shall find that the application of the principle of conservation to a certain cycle of reversible transformations directly expresses the principle of degradation of energy, and leads to the generalised analogue of the Second Law.

398. *General Law of Transformation.*—Let us denote the quantities which fix the physical condition of the system considered by the letters a, b, c, etc., while a', b', c', etc., denote very slightly

different values of the same quantities. The change of energy in any process is

$$E' - E = A(a' - a) + B(b' - b) + C(c' - c) + \ldots\ldots$$

where A, B, C, etc., are, in the ordinary language of dynamics, *forces* of the types which cause the changes $a' - a$, $b' - b$, $c' - c$, etc. They are the quantities upon the magnitude of which the tendency to transference, or transformation, of energy depends. Thus, if $A(a' - a)$ represents mechanical energy, A might represent pressure, while $a' - a$ represented displacement. In the case of thermal energy, A would represent temperature, while a' and a would represent entropy (§ 270). In the case of electrical energy, A and a would respectively represent potential and quantity of electricity, and so on.

Rankine was the first to introduce, into the discussion of transformation, this idea of two factors of energy, one of which is such that differences in its value determine the tendency to transference of energy. This factor might therefore be called the potential factor. Such a factor can always be found, whatever be the type of energy under consideration.

In the dynamical view, transformation is merely transference to a dissimilar system.

When the system returns, at the end of the process, to its original condition, we get

$$A(a' - a) + B(b' - b) + C(c' - c) + \ldots = 0 \ldots\ldots (1).$$

This equation is the expression of the principle of conservation, for it asserts that any amount of energy of the type (A, a), which enters the system, passes out of it again either in the form (A, a) or in some other form into which it is changed by the medium of the system, and we get finally $a' = a$, $b' = b$, etc.

If now we apply the principle to a closed reversible cycle of operations (§§ 264, 265), in which equilibrium is only disturbed to an infinitesimal extent at any stage, and in which, for example—1st, A' increases by $A - A'$, while a is constant; 2nd, a increases by $a' - a$, while A is constant; 3rd, A decreases by $A - A'$, while a' is constant; 4th, a' decreases by $a' - a$, while A' is constant, we get

$$(A - A')(a' - a) + (B - B')(b' - b) + (C - C')(c' - c) + \ldots = 0 \ldots (2)$$

where each quantity in brackets is so small that squares and products, etc., may be neglected.

In general, if the above cycle is strictly observed with regard to the energy (A, a), it cannot be simultaneously carried out for any other type. But the actual cycle for any other type may be supposed to be built up of an infinite series of such cycles on an infinitely smaller scale, so that the final effect is the same as if the cycle were carried out (*cf.* §§ 268-270).

Suppose that we are dealing with two types of energy only. In this case (2) becomes

$$(A-A')(a'-a)+(B-B')(b'-b)=0 \quad . \; . \; . \; . \; . \; . \; . \quad (3).$$

It is easy to see that equation (3) is in reality no less general than equation (2); for, before we could apply (2) to any particular case in which (say) energy of the type (A, a) is transformed into other types, we should have to break up the total quantity of energy of type (A, a) into its several parts, which are transformed, as far as possible, into the several types; and we should then have to consider each part separately. That is to say, we should have to resolve equation (2) into a series of equations of the form (3).

399. *Absolute Value of the Potential Factor.*—We may write (3) in the form

$$(A-A')(a'-a)=\beta'-\beta.$$

or

$$\frac{A-A'}{A}A(a'-a)=\beta'-\beta,$$

i.e.,

$$\frac{A-A'}{A}\alpha=\beta'-\beta, \quad . \; . \; . \; . \; . \; . \; . \quad (4),$$

where α is the amount of energy of the type (A, a) which enters the system in the first operation of the cycle described above, and $\beta'-\beta$ is the amount of the type (B, b) which is developed by transformation. The equation (4) expresses the principle of degradation, for it asserts that only the fraction $(A'-A)/A$ of the whole amount of energy of the type (A, a) which is supplied is changed into energy of the type (B, b). If α' is the amount of energy of the type (A, a) which remains untransformed, we may write (4) in the form

$$\frac{A-A'}{A}\alpha=\alpha-\alpha'.$$

This equation gives a definition of A on an absolute scale (which includes Kelvin's definition of absolute temperature as a special case), for a generalization of Carnot's reasoning regarding the efficiency of a reversible heat-engine shows that the working of this more general reversible system is independent of the nature of the working-substance. It may be put in the form

$$\Sigma\left(\frac{\alpha}{A}\right)=0,$$

which applies to a completed reversible cycle, and includes, as a particular case, the usual analytical expression of the Second Law of Thermodynamics. It is preferable, however, to work with the equivalent equation (3).

400. *General Relations.*— It will also be convenient now to change the notation, writing, say, a instead of $A-A'$, α instead of $a'-a$, b instead of $B-B'$, and β instead of $b'-b$. If, further, we regard, for the moment, a, α, b, β, as essentially positive quantities, we see that we must write (3) in the form

$$a\alpha \pm b\beta = 0 \quad . \; . \; . \; . \; . \; . \; . \; (4).$$

The sign which is to be taken has to be determined specially for each separate case of transformation. It must be remembered that the working-substance acts merely as a transforming medium, Energy is supplied to it in one form, and is given out from it in another. As a particular case, suppose that $a\alpha$, both factors being positive, represents energy supplied to the system. The positive, or negative, sign must be taken according as, b and β being positive, the product $b\beta$ represents energy taken from, or energy supplied to, the medium.

If (4) were written in the form $a\alpha \pm \beta b = 0$, we must take the opposite sign to that which we take when we write the second term in the order $b\beta$. For the cycle represented by βb is performed in the opposite order to that represented by $b\beta$.

Hence we get, when the ± sign is used in (4), the following relations:

$$\left(\frac{a}{\beta}\right)_{\alpha \text{ constant}} = \mp \left(\frac{b}{\alpha}\right)_{\beta \text{ constant}}$$

$$\left(\frac{a}{b}\right)_{\alpha \text{ constant}} = \pm \left(\frac{\beta}{\alpha}\right)_{b \text{ constant}}$$

A number of special examples are given in the immediately succeeding sections.

401. *Heat and Mechanical Work.*—Let θ, ϕ, p, v represent respectively small changes of temperature, entropy, pressure, and volume. The quantity $\theta\phi$ represents energy supplied to the system, while pv represents work done by it. Hence

$$\theta\phi + pv = 0,$$

giving
$$\frac{\theta}{v} = -\frac{p}{\phi}, \quad \frac{\theta}{p} = \frac{v}{\phi}, \quad \frac{\phi}{p} = -\frac{v}{\theta}, \quad \frac{\phi}{v} = \frac{p}{\theta},$$

four well-known thermodynamical relations.

If the quantity on the right-hand side of the first relation be multiplied and divided by Θ, the absolute temperature, the denominator represents the amount of heat which is supplied, at constant volume, in order to produce the variation, p, of pressure. If p and $\Theta\phi$ are positive, θ/v is essentially negative. Hence the equation asserts that *substances which, at constant volume, have their pressure raised (or diminished) by the application of heat, will fall (or rise) in temperature during adiabatic expansion; and the change of temperature, per unit change of volume, is numerically equal to the product of the absolute temperature into the change of pressure per unit of heat supplied.*

From the second relation, we see that, if θ and p are both positive, v and ϕ are necessarily of the same sign. Hence, multiplying and dividing the right-hand side by Θ, we see that *substances which expand (or contract) under constant pressure, when heat is supplied to them, rise (or fall) in temperature when they are subjected to adiabatic compression; and the change of temperature, per unit increase of pressure, is equal to the product of the increase of volume at constant pressure, per unit of heat supplied, into the absolute temperature.*

The third relation tells us that *substances which expand (or contract) under constant pressure when their temperature is raised emit (or absorb) heat in order that their temperature may remain constant when the pressure is increased; and the heat which is evolved (or absorbed) per unit increase of pressure is equal to the continued product of the temperature, the volume, and the expansibility.* For the expansibility is the quotient of v/θ by the volume.

Multiplying each side of the fourth relation by Θ, we see that *substances which absorb (or emit) heat when their volume increases isothermally, have their pressure, at constant volume, raised (or diminished) by increase of temperature; and the change of pressure, per unit rise of temperature, is equal to the ratio of the heat which is absorbed (or emitted), per unit increase of volume, to the absolute temperature.* If L represents latent heat, while v represents the change of volume of unit mass of the substance when it changes its state, the equation becomes

$$\frac{L}{\Theta v} \doteq \frac{p}{\theta},$$

from which we can calculate the change of the melting-point, or the boiling-point, which results from a given change of pressure.

402. *Latent Heat and Work during Evaporation.*—Let l be a small change in the amount of liquid, and let ρ be the density of

the liquid, while σ is the density of the vapour, and λ is the latent heat. From $\theta\phi+pv=0$ we get

$$\theta=p\frac{v}{\phi}.$$

Now $$v=\left(\frac{1}{\sigma}-\frac{1}{\rho}\right)l,$$

whence $$\theta=p\frac{\rho-\sigma}{\rho\sigma}\frac{l}{\phi}.$$

But $$\lambda l=\Theta\phi,$$

so $$=\frac{p\Theta}{\lambda}\frac{\rho-\sigma}{\rho\sigma},$$

which gives the change of the boiling-point dependent on the change p of pressure. The result applies also to the change in the melting-point, ρ being the density of the solid.

403. *Heat and Electrical Energy.*—Let I and i represent respectively small changes of electromotive force and quantity of electricity. Then Ii and $\theta\phi$ both represent energy supplied to the system, so that $\mathrm{I}i-\theta\phi=0$, whence, Θ being the actual temperature,

$$\frac{\Theta\phi}{i}=-\Theta\frac{\mathrm{I}}{\theta}$$

represents the amount of heat which must be withdrawn from the system (say a voltaic cell), during the passage of unit quantity of electricity, in order that the temperature may remain constant. This is Helmholtz's well-known expression.

If I represents the electromotive force in a thermo-electric circuit, we get

$$\frac{\mathrm{I}}{\theta}=\frac{1}{\Theta}\frac{\Theta\phi}{i}=\frac{\pi}{\Theta},$$

where π is the heat absorbed at the junction when unit quantity of electricity passes across it at temperature θ.

404. *Electrical Energy and Work.*—Consider a sphere of capacity C charged to potential V. The charge is CV. Hence, if changes of V and CV be represented by (V) and (CV), we get $(\mathrm{V})(\mathrm{CV})+pv=0$, and

$$p=(\mathrm{V})\frac{(\mathrm{CV})}{v}=\mathrm{V}(\mathrm{V})\frac{(\mathrm{C})}{\mathrm{V}},$$

since (§ 400) we have to take the value of $(\mathrm{CV})/v$ when V is constant. Now (§ 282) $v=4\pi\mathrm{C}^2(\mathrm{C})$, whence

$$p=\mathrm{V}(\mathrm{V})\frac{1}{4\pi\mathrm{C}^2}.$$

If Q be the total charge of electricity, while q is the change in its value, this gives

$$p=\frac{Qq}{4\pi C^4}.$$

Hence, by summation, we get the known expression for P, the pressure due to the total charge Q,

$$P=\frac{1}{8\pi}\frac{Q^2}{C^4}=\frac{R^2}{8\pi},$$

where R is the resultant electric force at the surface of the sphere. If we suppose electric energy to be stored in the medium, we see that $Q^2/8\pi C^4$ or $R^2/8\pi$ is also the expression for the energy per unit of volume at the surface.

405. *Electrification and Vapour Pressure.*—Let E be the electrical energy contained in a liquid sphere of radius C and density ρ, and let E and C increase by small amounts e and c respectively. We have

$$\frac{e}{c}=-\frac{1}{2}\frac{Q^2}{C^2}.$$

If m represents a small increase of the mass of the vapour, we have

$$m=-4\pi C^2\rho c.$$

Also we have

$$e=\frac{e}{c}\,\frac{c}{m}m=\frac{Q^2}{8\pi C^4\rho}m=pv.$$

But $$v=-\left(\frac{1}{\sigma}-\frac{1}{\rho}\right)m,$$

σ being the vapour density. Hence

$$p=-\frac{1}{8\pi}\,\frac{\sigma}{\rho-\sigma}\,\frac{Q^2}{C^4},$$

which agrees with J. J. Thomson's result.

406. *Surface Tension and Vapour Pressure.*—Let T and S represent respectively the surface tension and the surface of a drop of liquid, and let the other quantities be indicated as in the previous case. We get $ts-pv=0$.

Now $$s=8\pi Cc=-\frac{2m}{C\rho}=\frac{1}{C}\,\frac{2\sigma v}{\rho-\sigma}.$$

Hence $$p=2\frac{t}{C}\,\frac{\sigma}{\rho-\sigma},$$

or, summing from $T=0$ to $T=T$,

$$P-P_0=2\frac{T}{C}\,\frac{\sigma}{\rho-\sigma},$$

where P and P_e represent respectively the equilibrium-pressure of he vapour in contact with a plane surface and a spherical surface of radius C. This result was first given by Lord Kelvin.

407. *Induced Electromotive Forces.*—Consider two circuits in which electric currents of intensities C_1 and C_2 respectively circulate similarly. Let the coefficient of mutual induction be M, and let us take as an experimental result the expression

$$F = C_1 \frac{(C_2M)}{x}$$

for the force F which produces a small displacement x, in the time τ, brackets being used, as in § 404, to indicate a small change of the quantity within them. If E_1 and E_2 are the electromotive forces acting in the circuits, we have

$$Fx = C_1(C_2M) = e_1C_1\tau$$

$$e_1 = \frac{(C_2M)}{\tau}.$$

Similarly, $$e_2 = \frac{(C_1M)}{\tau}.$$

408. *Surface Tension and Solubility.*—If the energy due to the tension increases by the amount (TS), we have $(TS) - \theta\phi = 0$. Let a represent the increase in the amount of dissolved salt. If Θ be the temperature, $\Theta\phi$ represents the heat which is developed by the solution of the quantity a of the salt. Let h be the amount of heat developed per unit amount of salt dissolved, and we get $(TS)/a = h\theta/\Theta$—an equation of the type (4), § 399. Hence,

$$\theta = \frac{\Theta}{h} \frac{(TS)}{a}.$$

Therefore, if heat be developed, and if TS increases, during solution of the salt, the surface tension has the same effect as an increase of temperature. It will therefore increase the solubility if increase of temperature does so. If heat be absorbed, the reverse effect recurs.

409. *Pressure and Solubility.*— From the equation $\theta\phi + pv = 0$, or $\theta\phi - vp = 0$, we get the result $\theta/p = v/\phi$. From this, using a, h, Θ, as above, we find

$$\theta = p \frac{\Theta}{h} \frac{v}{a}.$$

Therefore, if h be positive while v increases as a increases, we see that increase of pressure causes an increase of solubility if increase of temperature does so.

410. *Relation between the Specific Heats of a Perfect Gas.*—It is very easy to prove, by means of thermodynamical considerations, a simple relation between the specific heats of a perfect gas. Let k

and c represent respectively the values of the specific heats at constant pressure and at constant volume, while R is the constant in the well-known expression $pv = Rt$. On the kinetic theory, the same amount of work is spent internally in raising the gas one degree in temperature, whether the volume be kept constant or not. Hence the excess of the specific heat at constant pressure over the specific heat at constant volume is the equivalent of the work spent in increasing the volume at constant pressure. Now, at the same pressure as before, but at a different temperature, t', we have $pv' = Rt'$ Hence, when $t' - t = 1$, we get $p(v' - v) = R$. Therefore R is the amount of work under consideration, and we get

$$k - c = R.$$

411. *Determination of the Absolute Scale of Temperature.*—When air is compressed by the sudden application of pressure, the heat developed is almost precisely equivalent to the work which is spent in producing the compression. Joule proved this by enclosing air in a strong vessel which could be placed in communication with another vessel, of the same size, which had been exhausted of air. Both vessels were placed in a large mass of water the temperature of which was accurately determined. When a stopcock in a tube connecting the two vessels was opened, the air rushed from the one vessel into the other so as to equalise the pressure throughout. The temperature of the vessel containing the expanding air was lowered, for work had been done during expansion, so that the air was cooled. That of the other vessel rose, for the violent impact of the air which rushed into it caused the development of heat. But the amount of heat which was absorbed in the one case was almost precisely equal to that which was evolved in the other, for the surrounding water, which was well stirred, showed no appreciable change of temperature.

A more accurate form of this experiment was subsequently adopted by Joule and Kelvin in their researches on the thermodynamical properties of gases. The gas under investigation was made to pass very slowly through a tube, in which a plug of cotton-wool was placed, and its pressure and temperature on both sides of the plug were observed.

It is easy to deduce a simple equation connecting the changes of temperature and pressure with the volume, the absolute temperature, and the expansibility of the substance. The expansibility may then be expressed, by means of Charles' Law, in terms of the temperature on the Centigrade scale, if the range of temperature be so small that the Centigrade and the absolute degrees are practically equal throughout its extent; so that the equation gives a direct comparison between the absolute and the Centigrade scales.

Boyles' and Charles' Laws give (§ 240) for a perfect gas $T = t + 1/a$, where T is absolute temperature. The investigation just alluded to gives

$$T = t + \frac{1}{a} - \psi,$$

where ψ, as Kelvin's and Joule's experiments indicate, is, in true gases, a small quantity—generally positive.

The experiments showed that all the true gases except hydrogen were made colder by their passage through the plug, and indicated that the absolute zero is about 273°·7 C.

412. *Basis of the General Law of Transformation.*—The truth of the Second Law of Thermodynamics (§ 266) rests entirely on the immensity of the number of particles contained in any portion of matter which is of a size comparable with the dimensions of our instruments and machines. An ordinary thermometer, placed in any position in a mass of air, might indicate uniformity of temperature; while another thermometer, sufficiently small in size and sufficiently delicate, might (rather, *would*) indicate rapid variations of temperature, and might even show that heat was passing from cold parts to hot parts of the given mass. For, the quickly moving molecules might occupy on the whole one portion of a volume so small as to contain only a few molecules, while the slowly moving molecules occupied the remainder, and some of the slowest of the quickly moving molecules might be exchanged for such of the slowly moving molecules as were actually moving more quickly than they were.

An *average* uniformity is preserved on the large scale, though, on a sufficiently small scale, it does not obtain. It is because of this average uniformity that the statement—that a heat engine cannot continually draw the heat which it transforms into work from a body colder than its condenser—which Kelvin made the basis of the Second Law of Thermodynamics is true. If we could control molecular motions, the law could be violated.

And, similarly, if complete control over the ultimate mechanism of transformation were possessed, the general law of transformation for all forms of energy could be violated.

CHAPTER XXXII.

THE ETHER.

413. **Whenever** mutual action is observed between two systems, it is possible to explain the accompanying phenomena, in great part at least, on either of two assumptions. We may assume that the action occurs directly at a distance, or we may assume that it is propagated by means of a material medium. But when we can show that the action takes time to travel from one point of space to another, the latter assumption only is tenable. For example, we have no evidence as yet (§ 76) that gravitational action is not instantaneously propagated, and therefore either assumption is valid; but the experiments of Hertz have shown that electromagnetic action requires a finite time for its propagation over a finite space, and therefore we must assume the existence of a medium by means of which that action is transferred.

Such a medium is termed an 'ether,' and we may therefore define the ether as a substance, other than ordinary matter, through which action is propagated.

At one time many ethers were supposed to exist—indeed, a new ether was postulated for the explanation of almost every new class of phenomena which presented itself. An ether was assumed in order to explain gravitation. Newton introduced a medium to account for the production of the 'fits' of easy reflection and transmission of luminous corpuscles, which he had to postulate for the explanation of the colours of thin plates, etc., on the corpuscular theory. The aid of another was invoked for the explanation of physiological phenomena, and so on—the attributes with which each was endowed being specially chosen in order to make it suit each particular case.

Such a procedure is totally unscientific, and it is now the aim of scientists to ascribe all actions, which apparently occur directly at a distance, to the intervention of a single medium. Of all the host of mediæval ethers, one alone remains—the medium whose existence was postulated by Huyghens in his explanation of the phenomena of light; and one of the great merits of Maxwell's modern electromagnetic ether is that it explains the phenomena of light as well as the phenomena of electricity and magnetism.

414. It is more easy to say what the ether is *not* than to say what it is. It is not a gas, like air; for transverse oscillations, such as those which take place in the propagation of light, die out with extreme rapidity in such a medium. Neither, for a like reason, is it a liquid like water. So far as this effect is concerned, it might be a transparent solid, for solids can transmit transverse oscillations. But, on the other hand, the rate at which transparent solids transmit such vibrations is immensely slower than the rate at which the ether transmits light. Hence the ether cannot be an ordinary transparent solid, although it interpenetrates such solids, and is hampered in its action by them—a fact shown by the diminished speed of light when passing through them.

Yet the ether, although it is not ordinary matter, must be material, *i.e.*, must possess inertia, for it transmits energy at a finite rate. And, for the same reason, it must possess rigidity and must be elastic. The rate of vibration of the parts of the medium is (*cf.* § 143) directly proportional to the square root of the rigidity, and is inversely proportional to the square root of the density. When red light passes through the ether, the rate of vibration is about 400,000,000 times per second. A steel tuning-fork which emits even the highest audible note would require to be immensely more rigid that it is if it were to vibrate at that rate; if its rigidity remained constant, it would require to be far less massive than it is. It would appear from Kelvin's calculations, based on a very plausible assumption, that the ether is about $(10)^9$ times less rigid than steel is; but, on the other hand, its density would appear to be about $(10)^{19}$ times less than that of steel.

415. It seems, therefore, that the ether acts as if it were an elastic solid. And yet the earth, in its course round the sun, moves through it without being subjected to any appreciable resistance; and the light which comes to us from a distant star gives no evidence of the disturbing effect which the earth might be supposed to have upon the ether in its neighbourhood as it moves quickly through it.

Young therefore suggested that the structure of the earth and other solid matter is such that the ether flows freely through it, being subjected to no alteration other than a change of density. But Stokes has shown that this rather startling assumption is probably unnecessary, since the speeds of the earth and of the particles of air in its atmosphere are small in comparison with that of light.

Of course, the free motion of the earth through the ether shows that, relatively to the moving earth, the ether acts like a practically non-viscous fluid, and we have to explain how it is that the ether can act both like a fluid and like an elastic solid. No explanation of this point can be more lucid than the original explanation given by Stokes: 'The plasticity of lead is greater than that of iron or copper, and, as appears from experiment, its elasticity is less. On the whole it is probable that the greater the plasticity of a substance, the less its elasticity, and *vice versâ*, although this rule is probably far from being without exception. When the plasticity of the substance is still further increased, and its elasticity diminished, it passes into a viscous fluid. There seems no line of demarcation between a solid and a viscous fluid. In fact, the practical distinction between these two classes of bodies seems to depend on the intensity of the *extraneous* force of gravity, compared with the intensity of the forces by which the parts of the substance are held together. Thus, what on the Earth is a soft solid might, if carried to the Sun, and retained at the same temperature, be a viscous fluid, the force of gravity at the surface of the Sun being sufficient to make the substance spread out and become level at the top; while what on the Earth is a viscous fluid might on the surface of Pallas be a soft solid. The gradation of viscous into what are called perfect fluids seems to present as little abruptness as that of solids into viscous fluids; and some experiments which have been made on the sudden conversion of water and ether into vapour, when enclosed in strong vessels and exposed to high temperatures, go towards breaking down the distinction between liquids and gases.

'According to the law of continuity, then, we should expect the property of elasticity to run through the whole series, only it may become insensible, or else may be mastered, by some other more conspicuous property. It must be remembered that the elasticity here spoken of is that which consists in the tangential force called into action by a displacement of continuous sliding; the displacements also which will be spoken of in this paragraph must be understood to be such displacements as are independent of condensation or rarefaction. Now, the distinguishing property of fluids is the extreme mobility of their parts. According to the views explained in this article, this mobility is merely an extremely great plasticity, so that a fluid admits of a *finite*, but exceedingly small amount of constraint before it will be relieved from its state of tension by its molecules assuming new positions of equilibrium. Consequently the same oblique pressures can be called into action in a fluid as in a solid, provided the amount of relative displacement of the parts be exceedingly small. All we know for certain is that the effect of elasticity in fluids [elasticity of form] is quite insensible in cases of equilibrium, and it is probably insensible in all ordinary cases of fluid motion. . . . But a little consideration will show that the property of elasticity may be quite insensible in ordinary cases of fluid motion, and may yet be that on which the phenomena of light entirely depend. When we find a vibrating string, the small extent of vibration of which can be actually seen, filling a whole room with sound, and remember how rapidly the intensity of the vibrations of the air must diminish as the distance from the string increases, we may easily conceive how small in general must be the amount of the relative motion of adjacent particles of air in the case of sound. Now, the extent of the vibration of the ether in the case of light may be as small, compared with the length of a wave of light, as that of the air is compared with the length of a wave of sound; we have no reason to suppose it otherwise. When we remember, then, that the length of a wave of sound in air varies from some inches to several feet, while the greatest length of a wave of light is about ·00003 of an inch, it is easy to imagine that the *relative* displacement of the particles of ether may be so small as not to reach, nor even come near to, the greatest relative displacement which could exist without the molecules of the medium assuming a new position of equilibrium, or, to keep clear of the idea of molecules, without the medium assuming a new arrangement which might be permanent.'

In this connection Kelvin refers to the properties of shoemakers' wax, which is so brittle that it will splinter under a sudden blow, and which will flow like a liquid into all the crevices of the vessel which contains it, while leaden bullets will sink down through it, and corks will float up through it—if only sufficient time be allowed (§ 61). The resistance to the passage of a bullet or a cork through it becomes smaller and smaller, the slower the motion becomes; and it may be that the motion of the earth through the ether is far less, relatively to the resisting power of the ether, than is the motion of the bullet or the cork relatively to the resisting power of the wax.

416. The above theory of the ether is known as the elastic-solid theory. This elastic solid cannot possess positive compressibility, for in that case a condensational-rarefactional wave—of whose existence we have no experimental evidence—might be propagated through it with finite speed. Hence Green, who investigated the properties of this ether, assumed that it was incompressible. He recognised the case of negative compressibility, but dismissed it with

the assertion that a medium which possesses negative compressibility—*i.e.*, a medium which expands when subjected to increased pressure, and contracts when pressure is removed—is necessarily unstable.

In order, on this theory, to account for the reflection and refraction of light at an interface, Green assumed that the ether had the same rigidity, but was of unequal density, on the two sides of the interface. This gave Fresnel's law in the case of vibrations perpendicular to the plane of reflection; but, in the case of vibrations in the plane of reflection, it gave a result which only coincided with Fresnel's law when the refractive indices of the two media were practically identical.

Lord Kelvin assumed that the ether consists of an inviscid fluid permeating the pores of an incompressible sponge-like solid: but the result deviated further from Fresnel's law than Green's did. He therefore had to abandon the doctrine of incompressibility; and, having pointed out that Green's negatively compressible medium was not unstable if it were infinite or had rigid boundaries, he assumed such negative compressibility as to make the velocity of the condensational-rarefactional wave zero. He assumed (like Green) equal rigidities of the ether in the media; but this condition has been shown to be necessary for stability when the other assumed conditions hold.

This contractile ether gives Fresnel's laws; and Glazebrook has shown that it explains the reflection and refraction of light by transparent bodies and by metals, double refraction and dispersion (including anomalous dispersion), and that it gives the correct expression for the velocity of light in a moving medium.

Kelvin has also shown how a model of a medium might be constructed by rigid jointed connections, and rigid revolving fly-wheels (or gyrostats in which a frictionless fluid circulated irrotationally) which has no intrinsic rigidity, *i.e.*, no intrinsic elastic resistance to change of shape, but which has a quasi-rigidity due to inherent resistance to rotation; which is absolutely devoid of resistance to change of volume or to irrotational change of shape; which therefore is incapable of transmitting condensational-rarefactional waves, but which can transmit vibrations like those of light. It is therefore a practical realisation of his contractile ether.

417. The electromagnetic theory of the ether has been discussed in the last chapter. It readily explains all the difficulties which originally beset the elastic-solid theory; and the question of the condensational-rarefactional wave never arises, for its velocity of propagation is infinite. On some points its results differ from those

of Kelvin's theory, but the differences are too small to admit of crucial tests being based upon them.

It must be observed, however, that this theory stands upon a different footing from the former. No fundamental assumption is made regarding the medium other than that it shall account for certain electrical and magnetic actions.

418. The most recent dynamical theory of the ether is one, elaborated by Larmor, realising MacCullagh's optical theory, in which all optical phenomena, exclusive of those involving absorption, are deduced from a general expression for the potential energy. The medium is supposed to be rotationally elastic, and to be incapable of resisting translational deformation. Centres of rotational strain existing in it have the permanence of vortices in a perfect fluid. The motion of these centres, or 'electrons,' as they are called, constitute electric currents. The known laws of inductive action and of electro-dynamical action are readily deduced, as also are the phenomena of radiation—including dispersion, metallic reflection, and the effect of motion of the material medium.

In certain respects the theory differs in its results from the results deduced from Maxwell's electro-magnetic theory, though the points of difference are not such that they can be used as experimental tests.

Material phenomena—including gravitation, inertia, and radiation from matter—are taken account of by regarding the electrons as the basis of matter.

419. The property of *dilatancy* in a medium composed of rigid particles in contact accounts for a number of natural phenomena and presents analogies to many others.

Let us suppose that we have a space filled with marbles or shot, each being in contact with another on various sides. This condition can be satisfied by different arrangements of the spheres, so that in some arrangements the volume occupied by a given number is less than the volume occupied by the same number in other arrangements. There is an arrangement of maximum volume and an arrangement of minimum volume, and we cannot change the hard spheres from one arrangement to another without altering the volume. (This explains the meaning of the term 'dilatancy.') Change of shape of such a mass of spheres cannot occur without simultaneous change of bulk. Hence, if the mass fill an inextensible, but flexible, boundary, no change of shape which necessitates change of volume can occur.

If the mass of spheres is enclosed by a smooth boundary, motion of the layer next the boundary will cause less alteration of volume than does the motion of a layer in the interior of the mass. Hence, when certain stresses are applied, there may be a streaming motion of the spheres along the boundary, while the rest of the spheres do not move. This conduction of the parts of the medium along a smooth surface resembles the conduction of electricity.

If, in a large mass of spheres in the condition of maximum density enclosed in an elastic boundary, one sphere grows in size, the whole medium at first undergoes dilatation. Then the layer next the growing sphere reaches the condition of maximum volume. After that, the layer next the sphere will be returning to the condition of minimum volume, while a layer a little farther out is at maximum volume. Later, there will be a succession of maxima and minima in the neighbourhood of the growing body.

When two bodies, growing in size, are present in the medium at a considerable distance apart, the resultant dilatation, at any point, due to both is less than the sum of the separate dilatations at that point due to each. Thus there will be a force of attraction between the bodies whose magnitude depends upon the rate at which the dilatation varies with the distance between the bodies. The dilatation becomes periodic when the bodies are near to each other, and attraction and repulsion occur alternately. This resembles the phenomena of molecular forces.

Instead of supposing that the boundary is elastic, we may assume that it is rigid, and that the growing spheres are elastic. We may even suppose that the spheres are rigid provided that the medium is composed of large spheres scattered uniformly among small spheres, for such a medium may possess elasticity in virtue of the propagation of distortional waves through it—just as a slack chain possesses elasticity when lateral vibrations are passing along it. Even if the small grains were at maximum density, distortional waves could pass, the distortions of the two sets of grains being opposite. This may throw light on electrodynamic and magnetic phenomena. Also, the separation of the two sets of grains would produce phenomena analogous to those depending on the separation of positive and negative electricities. And, with a certain arrangement of the large and small grains, the state of stress in the medium is the same as that which must exist in the ether in order to account for gravity.

Such, in brief outline, is Osborne Reynolds' theory of a granular ether.

420. It has been shown also that an ether consisting of vortices in a perfect fluid might be capable of transmitting light. And the instantaneous propagation of gravitational action (if it be instan-

taneous) does not in this case present so great difficulty; for, in a certain sense, each vortex occupies all the space which the fluid fills—its action is instantaneously felt in all parts.

421. It is not to be supposed that all these theories of the ether are necessarily antagonistic. The vortex theory, for example, may ultimately be the same as the elastic-solid theory.

CHAPTER XXXIII.

THE BASIS OF PHYSICAL BELIEFS AND THE ADVANCEMENT OF PHYSICAL SCIENCE.

422. The statement that science takes no cognizance of the question *how*, but only of the question *what*, is a statement which is very commonly made. And yet the reverse statement is much more nearly true.

We can only speculate as to the nature of matter, which is one of the most fundamental things in the physical universe. And, on the other hand, the ultimate object of the great bulk of experimental work is the discovery of relations amongst physical quantities. These relations then enable us to tell what changes we must directly produce in order that we may indirectly produce certain other changes. Thus, all such relations definitely tell us *how* certain results are to be brought about, although we may be in absolute ignorance as to the real nature of the quantities involved.

No doubt the discovery of *facts* or *phenomena* is the *first* object of investigation. But a mere collection of facts does not constitute a science.

423. The question with which we are at present concerned is the ground of our belief in the reality of phenomena and the ground of our belief in the relations, or laws, which are observed to regulate the occurrence of phenomena.

Belief in phenomena as objective realities is not a belief which can be arrived at by a process of reasoning. The physicist can give no reply to the man who asserts that there is no such thing as objective existence—that all phenomena are attributes of mind—other than the general one that such a statement is an assumption as great as the assumption which he makes himself. Yet he may add that his way of putting the matter is the one which appeals directly to the senses through which alone we get cognizance of the universe—a statement which is not affected by any question as to whether the senses and the universe are attributes of mind or not. If mind alone exists, he is quite willing to investigate mind in the best way which he has at his disposal.

The point at issue is the more directly important one—whether or not we may be entirely deceived regarding a phenomenon or a law connecting phenomena.

In § 4, it was stated that *conservation* is adopted as the great test of reality, and that this test led to the recognition of two classes of real things—matter and energy. Yet it is easy to see that this test may be delusive. For, if the linear magnitude of space and matter varied in any uniform way while the amount of matter remained constant, conservation of energy would still be satisfied when tested by our varying scales of length, while, with reference to an absolutely constant scale, it would not hold true.

The mirage is a temporary delusion. May not the desert and its heated air be a more permanent delusion? This is a question with which the physicist rightly refuses to deal until he finds evidence that such a consideration may be of value to him. He accepts as a reality any physical quantity which he can subject to investigation and which he can connect with other such realities by means of a determinate law. It is of no moment to him that his reality may only be an apparent one behind which the actual realities lie. He adopts the only means which he has of investigating the latter.

424. The one test which is ultimately adopted in physical science —as in all sciences—is the test of *fitness*. And that theory is most fit which is at once the most simple and the most comprehensive.

425. We have already found (§ 13) that all scientific knowledge is essentially based upon the results of observation and experiment, and it is therefore primarily by means of observation and experimental work that science is to be advanced. Every student can use and train his powers of observation, and almost every student, who so desires, can perform experimental work. Some of the greatest experimental results have been obtained by means of the simplest possible apparatus.

And, the experimental foundation having been laid, there is a wide field open to the theorist and the mathematical student. No doubt it does not fall to the lot of very many mathematical students to do pioneer work in science. But, as a very general rule, it is the case that the results obtained by the pioneers are not obtained in the simplest way possible, and thus much valuable work remains to be done in the simplification of present methods.

INDEX.

THE NUMBERS REFER TO THE SECTIONS.

ABERRATION, chromatic, 171
    spherical, 171
Absorption, co-efficient of, 179
Absorptive power, 177, 230
Acceleration, 26, 28, 29
Accumulator, water-dropping, 295
Achromatism, 172
Actinometer, 234
Activity, 50
Adiabatics, 269
Amplitude, 33
Analogy, 15
Atmolysis, 88
Atom, 66, 115-117
Avogadro's Law, 126

BEATS, musical, 150
Biprism, 188
Boiling, 249
Bolometer, 320
Boyle's Law, 79, 81, 125, 126
BREWSTER, 177, 182, 191, 197, 212, 217, 225, 237
Brittleness, 67

CALORIC, 228
Capillarity, 95-103
Carnot's cycle, 264, 265
CAUCHY, 122, 183, 232
Caustics, 159, 164
Cavendish experiment, 73
Charles' Law, 126, 239
Chemical combination, 254
Cohesion, 94, 113
Colour, 155
    blindness, 155
    body, 179
    complementary, 155
Colour, contrast, 155
    surface, 180
    vision, 155
    vision theories, 155
Colours of crystalline plates, 222, 223
    mixed plates, 196
    thick plates, 197
    thin plates, 192-195
Compressibility, 62
    of gases, 78
    of a perfect gas, 80
    of liquids, 89
    of solids, 104
    of vapours, 82
Conductivity, thermal, 127
Conservation, 4, 12
    of energy, 4, 10
    of matter, 4, 5
Consonance, 150
Contact, angle of, 97
Coronæ, 206
Coulomb's torsion balance, 280
Couple, 53
Critical angle of reflection, 163
Crucial test, 15
Crystalline structure, 119, 120
Curvature, 29

DENSITY, 46
    of earth, 73, 74
Deviation, minimum, 167
Dew, 251
Dialysis, 93
Diamagnetism, 348, 353
Dichroism, 179
Dielectrics, 274
Diffraction gratings, 207

Diffusion of gases, 85, 127
 of liquids, 92
Diffusivity, 85
Dilatancy, 419
Dilatation of gases, 239
 of liquids, 238
 of solids, 237
Dimensions, 18, 49
Discharging rod, 282
Dispersion, irrationality of, 172
 theories of, 183
Displacement, 24
Disruptive discharge, 294
Dissociation, 128, 254
Dissonance, 150, 151
Divisibility, 62, 115
Döppler's principle, 153
Dulong and Petit, law of, 244
Dynamical similarity, 49, 56, 59, 103, 143, 145, 146
Dynamics, 45
Dynamo, 373-76

Efficiency, thermo-dynamic, 267
Effusion, 86
Elasticity, 62
 of gases, 83
 of liquids, 90
 of solids, 110, 132
Electric absorption, 294
 charge, 278
 conductance, 310, 321
 current, 308, 309, 315, 320, 368
 density, 284
 discharge in gases, 323-41
 displacement, 292
 force, lines of, 286
 images, 285
 induction, 278, 292
 machines, 299
 potential, 281
 power, 383
 quantity, 278
 resistance, 309, 320, 384
Electrodynamometer, 320
Electrokinetic energy, 366
Electrolysis, 311-314
Electromagnetic induction, 363-365
 medium, 76, 394
 units, 385, 386
 waves, 395, 396
Electrometer, 298
Electromotive force, 281, 320
Electrophorus, 279
Electroscope, gold-leaf, 277
Electrostatic capacity, 282
 energy, 293
 units, 385
Emissivity, 177, 230
Empirical formulæ, 17
 laws, 17
Energy, 4, 7-11, 43
 conservation of, 10
 dissipation of, 11, 271, 272
 forms of, 8
 kinetic, 8
 potential, 8, 11
 transformation of, 9, 11
Entropy, 270, 272
Epoch, 33
Equation, personal, 16
Equilibrium, of fluids, 58
 of particles, 47
 of rigid bodies, 52
 stable, 15, 103, 304
Equipotential surfaces, 286
Eriometer, 206
Errors, instrumental, 16
 observational, 16
 probable, 16
Ether, 154, 176
Evaporation, 128
Exchanges, theory of, 229, 230
Extension, 62

Faraday, 250, 283, 292, 311, 363, 389
Fluid, equilibrium of, 58
 motion of, 42, 57
 motion of perfect, 57
Fluorescence, 182, 227
Focal lines, 159, 164
Focus, principal, 158, 168
Force, 43, 44
 lines and tubes of, 286
 measurement of, 46
 moment of, 53
Forces, composition of, 47
 molecular, 67, 101
 range of, 121
Form, 62
Freedom, degrees of, 25, 36, 38
Freezing mixtures, 253
Fresnel, 161, 186, 188, 211, 213, 214, 219
Fresnel's rhomb, 226
Friction, 44

Galvanometer, 320, 367

Gas battery, 318
Gaseous pressure, 125
Geysers, 249
Glaciers, motion of, 247
Graphical method, 17
Gravitation, 8, 44, 62, 69-77
  law of, 71
Gravity, centre of, 72
  hypotheses in explanation of, 75, 76
  specific, 70
GREEN, 219, 416
Gyration, radius of, 53
Gyrostats, 39, 132

HAIDINGER'S brushes, 223
Hall's effects, 391
Halos, 173
HAMILTON, Sir W. R., 220
Harmonic motion, simple, 33, 34
Harton experiment, 74
Heat, conduction of, 256, 261
  convection of, 256, 262
  latent, 246, 248, 250
  mechanical equivalent of, 263
  radiant, 227
  radiation of, 231-233, 235
  specific, 242, 243, 245, 410
  total, 250
HELMHOLTZ, Von, 148, 155, 183, 223, 312, 403
Hertz, 395
Hoarfrost, 251
Hodograph, 31
HOOKE, 111, 123, 211
Hooke's Law, 111
Hope's experiment, 238
HUYGHENS, 161, 208, 209, 211
Hygrometer, 251
Hypothesis, 15
Hysterēsis, 356-358

IMAGE, 157
  real, 158, 170
  virtual, 158, 170
Impact, 51
Impenetrability, 62
Impulse, 51
Indicator diagram, 268-271
Inductive capacity, specific, 283
Inertia, 6, 62
  centre of, 52, 73
  moment of, 53, 54
Isothermals, 269

JOULE, 123, 124, 263, 358, 411
Joule's Law, 319

KELVIN, 15, 76, 102, 113, 117, 122, 124, 131, 183, 234, 247, 266, 267, 305, 390, 406, 411
KEPLER'S laws, 71
Kerr's effects, 392
Kinetics, 45
KIRCHHOFF, 177, 178, 230
Kirchhoff's Laws, 310

LEAST action, 175
  confusion, circle of, 159
  time, 175
Lenses, 168, 169
  achromatic, 172
LE ROUX, 181, 300, 306
LE SAGE, 75, 123
Leyden jar, 282
Light, absorption of, 176-183
  colour of, 155
  diffraction of, 198-207
  dispersion of, 171, 172, 181, 183
  intensity of, 152
  interference of, 184-197, 221
  nature of, 154, 216
  polarisation of, 211-226
  propagation of, 152, 161, 199
  radiation of, 176-178
  reflection of, 156-161, 214
  refraction of, 162-175, 214
  refraction (double), 208-210, 218, 219, 224
  scattering of, 156
  speed of, 153
  theories of, 154, 160, 161, 174, 175
Lightning-rod, 295
LLOYD, 187, 221

MACCULLAGH, 213, 214, 217, 219
Magnetic axis, 343
  flux, 354
  force, lines of, 342
  induction, 349, 352
  intensity, 344
  moment, 343, 349
  permeability, 353, 355
  pole, 343, 349
  potential, 345, 346
  reluctance, 354
  retentiveness, 349
  shell, 344, 346

Magnetic susceptibility, 353, 355, 357, 358
  viscosity, 355, 356
Magnetism, residual, 349, 355
  terrestrial, 360
  theories of, 361
Magnetometer, 359
Magnetomotive force, 354
Magnets, coercive force in, 349, 355
  electro, 349, 372
  permanent, 349
Malleability, 67
Malus, 211, 212, 217
Mass, 7
  measurement of, 46
Matter, 4, 5
  definitions of, 60
  general properties of, 62
  heterogeneity of, 118, 121, 122
  special properties of, 63, 67
  specific properties of, 65
  states of, 61
Maxwell, 84, 121, 123, 128-130, 155, 224, 292, 293, 361, 390, 394
Metallic reflection, 180, 217
Microscope, 170
Mirage, 166
Molecules, 66, 130
  size of, 122
Moment, 32
Momentum, 46
Motion, laws of, 45-47, 50
  of centre of inertia, 52
  of connected particles, 50
  of non-rigid solids, 56
  of particles, 46-48
  of rigid body, 37, 38
Motor, electric, 380
Musical intervals, 142

Nebular hypothesis, 77
Neumann, 213, 214, 219
Newton, 45, 49, 50, 71, 72, 76, 161, 197, 211, 233
Newton's rings, 195
Note, 133

Ohm's Law, 310, 314
Osmose, 93

Paramagnetism, 348, 353
Paraselenæ, 173
Parhelia, 173
Path, mean free, 124, 129
Pendulum, 9
  ballistic, 51
Peltier effect, 304
Period, 33, 35
Phase, 33
Phosphorescence, 182, 227
Photometer, 152
Piezometer, 89
Polarising prisms, 226
Porosity, 62
Position, 23
Potential, 281
  gravitational, 281
Precession, 39
Prevost, 123, 177, 229
Primary cells, 316
Prisms, 167
Projectiles, 30
Pyrheliometer, 234
Pyro-electricity, 206
Pyrometers, 241

Quartz fibres, 73

Radiometer, 129
Radiomicrometer, 307
Rainbows, 173
Rankine, 228, 269, 390, 398
Reflection, total, 163
Refraction, conical, 220
  index of, 162, 197
Regnault, 238, 239, 243, 245, 249-251
Residual discharge, 294
Resonance, 148
Restitution, co-efficient of, 51
Rigidity, 62, 67, 104, 107, 132
  flexural, 109
  torsional, 107
Ripples, 59
Rotation, 36
Rotational equilibrium, 55

Scalar quantity, 24
Schehallion experiment, 74
Secondary cells, 318
Shear, 40
Solar radiation, 234
Solenoid, 347
Solution, 253
Sound, diffraction of, 139
  intensity of, 136
  interference of, 140, 150
  pitch of, 141
  propagation of, 134

Sound, quality of, 149
  reflection of, 137
  refraction of, 138
  speed of, 134, 135, 147
  wave-length of, 133
Spectrometer, 178
Spectrum analysis, 178
  diffraction, 207
  heat, 232
  refraction, 171
Speed, 25
  average, 25
  mean square, 124
Specific heat of electricity, 305, 306
Statics, 45
Stewart, Balfour, 177, 230, 233, 235
Stokes, 182, 191, 198, 199, 208, 213, 216, 219, 415
Strain, 40, 41
Stress, 50
Surface-tension, 96, 101-103
Syren, 141

Tait, 129, 252, 258, 303, 306
Telephone, 9
Telescope, astronomical, 170
Temperature, 236
  absolute, 267, 269, 411
  absolute zero of, 240, 411
  critical, 20, 252
  measurement of, 241
  triple point, 20
Tenacity, 67, 114
Theory, 15
  mathematical, 15
Thermal capacity, 242
  conductivity, 257, 258
Thermodynamic motivity, 272
Thermodynamics, first law of, 263
  second law of, 266
Thermo-electric diagram, 303, 306
  power, 301
Thermometer, 241
Thermometer, hypsometric, 249
  wet and dry bulb, 251
Thermometric conductivity, 258
  of crystals, 260
  of earth's crust, 259
Thermopile, 301, 307
Thomson effect, 305
Tone, 133
  combination, 151
  partial, 148
Transpiration, 87
Twist, 38

Vector, 24
Velocity, 25, 28
  angular, 29, 39
Vibrations of air-columns, 147
  of plates, 145
  of rods, 143, 144
  of strings, 146
Viscidity, 91
Viscosity, 62, 67
  of gases, 84, 127
  of liquids, 91
  of solids, 112
Voltaic batteries, 317
Vortex atoms, 117, 132
  motion, 42
Vortices, molecular, 228, 390

Water equivalent, 243
Wave-length, 35
  motion, 35
Waves, long or solitary, 59
  on stretched cords, 56
  oscillatory or free, 59
Weight, 46, 62, 69
Wheatstone's bridge, 320
Work, 7, 43

Young, 155, 185, 195, 196, 206, 211
Young's modulus, 106

Zone plates, 204

THE END.

Baillière, Tindall & Cox, 20 and 21, King William Street, Strand.